新编电气与电子信息类本科规划教材

信息论与纠错编码

（第2版）

孙丽华　陈荣伶　编著

電子工業出版社
Publishing House of Electronics Industry
北京・BEIJING

内 容 简 介

本书重点介绍了信息论与纠错编码的基础内容，全文共分 10 章。内容包括信息及信息的度量、离散信源及信源熵、离散信道及信道容量、信源编码定理和信道编码定理、平均失真测度和信息率失真函数、率失真编码定理、纠错编码代数基础、线性分组码、循环码和卷积码。

本书文字通顺，深入浅出，概念清晰，对一些较难理解的概念，辅有较多的例题，并配有免费电子课件、习题解答等教辅资料。

本书适合作为高等院校理工类本科电子技术、信息工程、通信工程、雷达、计算机、自动化、仪器仪表等相关专业的教材，也可作为信息科学及系统工程等专业教学人员及科研人员的参考书。

图书在版编目（CIP）数据

信息论与纠错编码/孙丽华，陈荣伶编著．—2 版．—北京：电子工业出版社，2009.8
新编电气与电子信息类本科规划教材
ISBN 978-7-121-09369-2

Ⅰ. 信…　Ⅱ.①孙…②陈…　Ⅲ. ①信息论—高等学校—教材②信源编码—编码理论—高等学校—教材③信道编码—编码理论—高等学校—教材　Ⅳ. TN911.2

中国版本图书馆 CIP 数据核字（2009）第 133392 号

策划编辑：王羽佳
责任编辑：秦淑灵
印　　刷：
装　　订：
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：787×1 092　1/16　印张：14.5　字数：371 千字
印　　次：2009 年 8 月第 1 次印刷
印　　数：4 000 册　　定价：25.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

信息论是应用近代概率统计方法研究信息传输、交换、存储和处理的一门学科，也是源于通信实践发展起来的一门新兴应用科学。当前人类已步入信息社会，随着信息概念的不断深化，信息在科学技术上的重要性也早已超越了狭义的通信工程的范畴，受到越来越多的关注。

在高等院校中，信息工程类专业是最热门的专业之一，信息技术已经改变了很多传统电子类专业的知识结构。在这种形式下，许多高校都在相关专业开设了信息论课程，作为本科生、研究生的必修或选修课程。一方面，这门课是电子类专业的核心课程，很多学校都把它作为必修课或必选课；但另一方面，这门课程涉及多门工程数学理论，例如概率统计论、线性代数、近世代数等，一直有教师难教、学生难学的说法。本书力求在内容筛选及编排上以读者最易接受的方式介绍信息理论的知识。

本书包括“信息论与纠错编码”的基本内容及应用。

第 1 章为“信息论基础”，介绍了信息论的基本概念，以及本书的研究对象——各种信源和信道。

第 2～6 章为信息论部分，介绍了信息的度量，内容主要围绕香农三大定理展开，研究在不允许失真的情况下信息传输率的极限值，以及给定信源并且允许一定失真的条件下信息速率的极限值，并研究在误码率小于给定值的条件下如何最有效地利用信道的传输能力。

第 7 章讲述了纠错编码所必需的数学知识。

第 8～10 章为纠错编码部分，纠错编码是后人沿着香农指明的可行方向，为寻求有效而可靠的编译码方法而发展起来的一门学科，主要研究在有噪信道条件下各种可行的编码方案及实施技术。

与现有的各种“信息论与编码”教材相比，本教材特色如下。

（1）本书力图在编排上由浅入深，循序渐进，希望读者以易于掌握的方式接受信息论与纠错编码方面的基本理论知识；

（2）对于部分具有结论性、指导性的定理，教材省去了冗长烦琐的定理证明，注重物理概念的阐述以及对后人工作的指导意义；

（3）增加了纠错编码部分的内容。教材第 8，9 章和第 10 章分别论述纠错码中最基本的线性分组码、循环码和卷积码的编译码理论，并列举了几种常用的码，如汉明码、BCH 码和卷积码，介绍了它们主要的编译码方法。

（4）对一些难以理解的概念，本书配有较多的例题，以帮助学生理解抽象定理。各章后面配有较多难易程度不等的思考题和习题，以供选用。

（5）在第 1 版的基础上，收集了使用该书师生的反馈意见，对一些较难理解的概念，增加了较多的例题、思考题和习题，并将第 1 版中第 5 章和第 6 章的内容做了一些调整，使之更趋合理。

本书全部内容约需 62 学时，不同专业可根据需要进行调整。

本书配有**教学课件和配套辅助文件**，需要者可到华信教育资源网 http://www.hxedu.com.cn **免费注册下载**。

本书第 1，7，10 章、附录 1 和附录 2 由陈荣伶编写，第 2～6 章、第 8 章和第 9 章由孙丽华编写，由孙丽华负责全书的策划、修改和统编。

本书在编著过程中参阅了一些国内外相关著作，这些著作已在参考文献中一一列出，在此谨向有关作者表示深深的谢意！

在此也向第 1 版的合作者谢仲华表示深深的谢意！

本书在编写过程中得到电子工业出版社的大力支持，王羽佳和秦淑灵编辑做了大量的工作，使本书得以顺利出版，在此一并表示衷心的感谢！

本书涉及知识领域广泛，知识变化日新月异，由于时间和水平的限制，难免有差错和不足之处，敬请读者指正！

孙丽华

2009 年 6 月

于南昌

目　录

第 1 章
信息论基础

内容提要

信息论是人们在长期的通信实践中发展起来的一门新兴应用科学，它应用近代概率统计方法来研究信息的传输、交换、存储和处理。自 Claude E.Shannon 发表《通信的数学理论》开始，随着信息概念的不断深化，信息理论在科学技术上的应用早已超越了狭义的通信工程范畴，受到越来越多的关注。本章首先引出信息、消息的概念，讨论信息论的研究范畴；然后简述数字通信系统的结构和特点，讨论离散信源、波形信源和离散信道的数学模型和特点；最后介绍几种常见的离散无记忆信源和离散无记忆信道。

知识要点

信息的定义、信息论的研究范畴、数字通信系统的组成、离散信源和离散信道的数学模型。

教学建议

本章是信息论的基本概念，后续信息的度量、信源编码及信道编码都是围绕本章的概念而展开的，建议学时数为 3 学时。

1.1 信息的概念

人们认为，物质、能量和信息是构成客观世界的三大要素。信息是物质和能量在空间和时间上分布的不均匀程度，或者说信息是关于事物运动的状态和规律。物质、能量和信息三者相辅相成，缺一不可。没有物质和能量就不存在事物的运动，也就没有运动状态和规律，当然也就没有信息；反过来，事物在运动，这种运动的状态和规律就成为信息。

因此可以说，没有了信息也就没有了一切！

那么，到底什么是信息呢？

信息有以下三种不同层次的定义。

- 语法信息：它是事物运动状态和规律的本身。它只研究事物运动可能出现的各种状态以及这些状态之间的关系，不涉及状态的含义和效用。
- 语义信息：它是事物运动状态和规律的具体含义。它研究各种状态和实体间的关系，即研究信息的具体含义。
- 语用信息：它是事物运动状态和规律及其含义对观察者的效用。它研究事物运动状态和规律与使用者的关系，即研究信息的使用价值。

语法信息是最抽象最基本的层次，通信工程中的信息传递问题正是基于语法信息。

我们知道，通信系统中形式上传输的是消息，消息与信息有何区别呢？

信息是一个抽象的概念，而消息是具体的，如一场足球赛事的状况，我们可以分别通过电视、广播和报纸来了解，这其中就涉及图像、语言、文字等不同形式的消息。可以说，消息是能被人们的感觉器官感知的客观物质和主观思维的运动状态（或存在状态）。在通信之前，收信者无法判断发信者将发送何种状态的消息。通过消息的传递，收信者知道了消息的具体内容，我们说收信者获得了信息。由此看来，通信系统中形式上传输的是消息，实质上传输的是信息。换句话说，消息中包含信息，消息是信息的载体，信息是消息中包含的有意义的内容。不同形式的消息可以包含相同的信息，如语言和文字发送同一天的天气预报，信息内容是相同的。

信息论是源于通信实践发展的一门新兴学科，从通信的角度讲，信息论是应用近代概率统计方法研究信息的基本性质及度量方法，研究信息的获取、传输、存储和处理的一般规律的科学。自 1948 年美国科学家香农（Claude E.Shannon）在 Bell 系统技术杂志上发表重要著作《通信的数学理论》开始，信息论就开启了迅猛发展的篇章。香农信息理论以概率论为工具，定量地描述了信息的含义，通过信源编码定理和信道编码定理指出，在通信系统中采用适当的编码后，能够在随机噪声干扰下有效而可靠地传送信息，并从理论上论证了信息传输的一些基本界限。随后这一理论引起了数学家和通信工作者的高度关注，陆续推出了 Fano 信源编码、Huffman 信源编码、Viterbi 信道译码、数据压缩、网络信息论等各种理论。

信息论是在长期的通信工程实践中发展起来的，但是，由于信息问题本身具有极为广泛的意义，信息论很快就渗透到其他领域并相继取得了新的发展，如信息生物学、量子密码学、信息经济学等。

对于信息论的研究，一般划分为以下三个不同的范畴。

- 狭义信息论：即通信的数学理论，主要研究狭义信息的度量方法，研究各种信源、信道的描述和信源、信道的编码定理。

- 实用信息论：研究信息传输和处理问题，也就是狭义信息论方法在调制解调、信息处理、检测与估计以及保密理论等领域的应用。
- 广义信息论：包括信息论在自然和社会中的新应用，如模式识别、机器翻译、自学习自组织系统、心理学、生物学、经济学、社会学等一切与信息问题有关的领域。

在信息时代，人们对于信息的理解远远超出了狭义信息论的讨论范围，要求进一步认识和发展信息概念和信息理论。信息科学的很多问题还在探索之中，本书只限于讨论在通信学科中已建立了完整理论并取得重大技术成就的狭义信息论。

1.2 数字通信系统

通信的基本问题是在彼时彼地精确地或近似地再现此时此地发出的消息。

消息分为两类：离散消息和连续消息。离散消息也称为数字消息，消息状态数是可数的或离散型的，如符号、文字等。连续消息又称为模拟消息，消息状态是连续变化的，如语音、图像等。通信中消息被转换成电信号，按信号特征的不同，通信系统分为数字通信系统和模拟通信系统。相比模拟通信系统，数字通信系统更能适应对通信技术的高要求，它具有以下优点：

① 抗干扰能力强，中继时可再生，可消除噪声积累；

② 差错可控制，可改善通信质量；

③ 便于加密和使用 DSP 技术处理；

④ 可综合传递各种消息，传送模拟消息时，只要在发送端增加模数转换器，在接收端增加数模转换器即可。

数字通信系统的主要性能指标如下。

（1）有效性

一般用码元传输速率或信息传输速率来衡量通信的有效性。码元传输速率是每秒钟传送的码元个数，单位为波特。信息传输速率是每秒钟传送的信息量，单位为比特/秒。在二元通信系统中，这两种速率在数值上是相等的。

（2）可靠性

用误码率或误信率表示。误码率是指错误接收码元在传送总码元数中所占的比例。误信率也称误比特率，指信息量被传错的概率。误码率和误比特率越低，可靠性越强。

各种数字通信系统，如电报、电视、遥控和雷达系统，虽然形式和用途各不相同，但从信息传输的角度来看，它们在本质上有许多共同之处。对有收、发两端的单向传输系统，一般可概括为图 1-1 所示的统计模型。

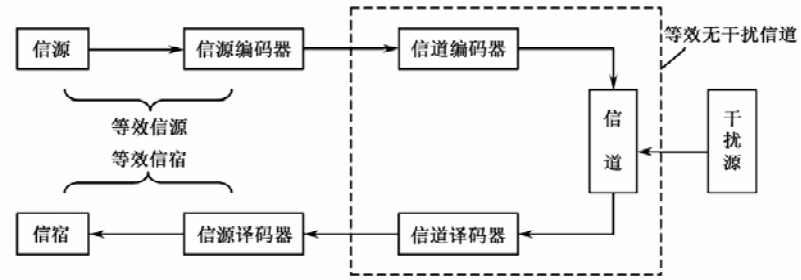


图 1-1 数字通信系统模型

这个模型包括以下五部分。

1．信源

信源是产生消息的源。消息可以有多种形式，如语言、文字、图像等。消息可以是离散的，也可以是连续的。消息中包含着信息，或者说消息是信息的载体，通信的结果是为了获得信息。信源输出的消息是随机的，在没有收到这些消息之前，收信者所获得的信息的多少是不确定的。在信息论中用随机变量或随机过程来描述消息。

2．编码器

编码器是将消息变换成适合于信道传送的信号的设备。编码器分为两种：信源编码器和信道编码器。

信源编码器对信源输出的消息进行适当的变换和处理，来提高信息传输的效率。信源编码常采取消除冗余或根据需要的质量标准去掉一些次要信息等措施。例如，在发中文电报时，需将每个汉字编码为 5 位等重码，为提高传输效率，可在不改变语意的情况下将词语或句子压缩。比如，“奥林匹克运动会”可压缩为“奥运会”，这样，原意不变，而冗余度大大减少。再如，电视信号只含 4%的有效信息，采用无失真压缩编码可达 30 倍的压缩率，而多媒体会议采用有损压缩则可压缩上百倍。

经过信源编码的码字序列均被认为是重要信息，如果在传输中受到干扰发生错误或误差超过给定标准，则不满足接收者的质量要求。信道编码是为了抵抗信道的干扰，提高通信的可靠性而对信源编码器的输出进行的变换和处理。为了提高可靠性，可以扩展带宽，降低传输速率，等等，而最常用、最有效的措施是采用差错控制编码，增加冗余码元来自动纠错或检错重发。如奇偶检验码，通过增加一位奇/偶检验位可检验出奇数位错，该奇/偶检验位与信息的内容无关，是个冗余码元。

3．信道

信道是信息传输和存储的媒介，如光纤、电缆、无线电波、磁盘、书籍等。信道上不可避免地存在各种干扰源，比如来源于无线发射机的无线电干扰、电气设备的工业干扰，以及宇宙射线的天电干扰及电子器件的内部干扰等。为了分析方便，我们将系统其他部分产生的各种干扰都等效地折合成信道干扰。信道的输出是信道输入信号和干扰的组合，由于干扰往往具有随机性，所以信道的特性也用一个随机过程来描述。

4．译码器

译码是编码的逆变换，分为信道译码和信源译码。由于信道干扰的影响，信道输出的信息序列中可能已有错误，信道译码就是从受干扰的信号中尽可能地纠正其中的错误，再现信源编码器的输出。

信源译码就是将信道中传输的各种信号还原成收信者能感知的消息。

5．信宿

信宿是消息的接收者。可以是人，也可以是机器。

通信的最终目的是有效地、可靠地传递消息。有效性和可靠性两者往往相互矛盾，要提高有效性，就要减少信源的冗余度，缩短每个数据码元所占的时间，这样势必使波形变窄，能量

减少，从而使受到干扰后产生错误的可能性增加，传送消息的可靠性降低；若要求可靠，就要增加纠错检错码元，这样增加了信道的冗余度，从而使传送消息的效率变慢。如上例中，若发电报“奥运会”，当我们收到电报“X 运会”时，无法判断所发电报是“奥运会”、“亚运会”还是“农运会”等，可见，所发电文虽然冗余度很小，但容错能力较差；而如果发电报“奥林匹克运动会”，当收到电报“X 林匹克运动会”时，我们很容易纠正电文的错误，译为“奥林匹克运动会”，说明信源的冗余度对于抵抗信道的干扰是有益的。那么怎样将矛盾的可靠性和有效性统筹兼顾？香农的信息理论指出，在一定的准则下，可以实现有效而可靠的通信。

数字通信系统的模型不是一成不变的，它根据实际情况而定。例如，在研究信息传输的有效性时，可将信道编码器、信道译码器和信道组合起来，等效为一个无干扰信道，这样信源编码器的研究只和信源、信宿有关；而在研究信息传输的可靠性时，可将信源译码器和信宿等效为信宿，将信源和信源编码器等效为一个对于信道编码器而言的信源，这样信道编码的研究只和信道有关，与信源、信宿无关。理论表明，这种简化方法对大多数理论结果没有太大限制。

1.3 信源及其数学模型

信源是产生消息的源，消息中含有信息。信息是抽象的，而消息是具体的，所以可通过消息来研究信源，研究信源各种可能的输出及输出各种可能消息的不确定性。虽然消息是随机的，但其取值服从一定的统计规律，因此信息论中用随机变量或随机过程来描述消息，或者说，用一个样本空间及其概率测度$\{X, q(X)\}$来描述信源。

根据样本空间 X 取值分布的不同情况，信源可分为以下类型。

- 离散信源：消息集 X 为离散集合。即时间和幅度取值都离散的信源。
- 连续信源：时间离散而幅度取值连续的信源，如温度、压力等。
- 波形信源：时间连续的信源，如语言、图像信源等。

连续信源和波形信源输出的消息都可以经过抽样和量化分别处理成时间和幅度取值都离散的消息，因此本书中主要讨论离散信源的情况。

根据信源的统计特性，信源又分为以下两种类型。

- 无记忆信源：X 各时刻的取值相互独立。
- 有记忆信源：X 各时刻的取值互相有关联，如中文句子中前后文字的出现是有依赖性的。有记忆信源的数学模型通常采用联合概率空间来描述。

1.3.1 离散无记忆信源

1. 离散无记忆信源

离散无记忆信源（Discrete Memoryless Source，DMS）输出的是单个符号的消息，不同时刻发出的符号之间彼此统计独立，而且符号集中的符号数目是有限的或可数的。离散无记忆信源的数学模型为离散型的概率空间，即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_k \\ q(x_1) & q(x_2) & \cdots & q(x_k) \end{bmatrix}$$

式中，$q(x_i)$是信源输出符号消息 x_i 的先验概率，满足 $0 \leqslant q(x_i) \leqslant 1$，$1 \leqslant i \leqslant k$，且 $\sum_{i=1}^{k} q(x_i) = 1$。

信源每次输出的符号消息 $x_i \in \{a_1, a_2, \cdots, a_k\}$，即 x_i 的取值必定是 k 个符号$\{a_1, a_2, \cdots, a_k\}$中的某一个。

【例 1.1】 二进制对称信源只能输出符号 0 或 1，输出 0 的概率为 p，输出 1 的概率为 $1-p$，其概率空间可描述为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$$

【例 1.2】 随机掷一个无偏骰子，可能出现的点数与其概率分布为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

2. 离散无记忆扩展信源

在实际情况下，信源输出的消息往往不是单个符号，而是由许多不同时刻发出的符号所组成的符号序列。设序列由 N 个符号组成，若这 N 个符号取自同一符号集$\{a_1, a_2, \cdots, a_k\}$，并且先后发出的符号彼此间统计独立，则我们将这样的信源称为离散无记忆的 N 维扩展信源，其数学模型为 N 维概率空间，即

$$\begin{bmatrix} \boldsymbol{X} \\ q(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{k^N} \\ q(\boldsymbol{x}_1) & q(\boldsymbol{x}_2) & \cdots & q(\boldsymbol{x}_{k^N}) \end{bmatrix}$$

式中，$\boldsymbol{x}$ 为各种长为 N 的符号序列，$\boldsymbol{x} = x_1 x_2 \cdots x_N$，$x_i \in \{a_1, a_2, \cdots, a_k\}$，$1 \leqslant i \leqslant N$，序列集 $\boldsymbol{X} = \{a_1 a_1 \cdots a_1, a_1 a_1 \cdots a_2, \cdots, a_k a_k \cdots a_k\}$，共有 k^N 种序列，$\boldsymbol{x} \in \boldsymbol{X}$。由于序列中前后符号无关，故序列的概率 $q(\boldsymbol{x}) = q(x_1 x_2 \cdots x_N) = \prod_{i=1}^{N} q(x_i)$，说明符号序列的概率是序列中各个符号概率的乘积，满足 $0 \leqslant q(\boldsymbol{x}_i) \leqslant 1$，$1 \leqslant i \leqslant k^N$，且 $\sum_{X} q(\boldsymbol{x}) = 1$。

【例 1.3】 将二进制对称信源进行三维扩展，则信源序列共有 2^3 种：000，001，010，011，100，101，110，111。由 $q(0) = p$，$q(1) = 1-p$ 可得序列的概率依次为

$$q(000) = q(0) \cdot q(0) \cdot q(0) = p^3 \qquad q(001) = q(0) \cdot q(0) \cdot q(1) = p^2(1-p)$$
$$q(010) = q(0) \cdot q(1) \cdot q(0) = p^2(1-p) \qquad q(011) = q(0) \cdot q(1) \cdot q(1) = p(1-p)^2$$
$$q(100) = q(1) \cdot q(0) \cdot q(0) = p^2(1-p) \qquad q(101) = q(1) \cdot q(0) \cdot q(1) = p(1-p)^2$$
$$q(110) = q(1) \cdot q(1) \cdot q(0) = p(1-p)^2 \qquad q(111) = q(1) \cdot q(1) \cdot q(1) = (1-p)^3$$

将这 8 种序列看成 8 个消息，得到一个新的信源，即

$$\begin{bmatrix} \boldsymbol{X} \\ q(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ p^3 & p^2(1-p) & p^2(1-p) & p(1-p)^2 & p^2(1-p) & p(1-p)^2 & p(1-p)^2 & (1-p)^3 \end{bmatrix}$$

1.3.2 离散有记忆信源

汉字或英文字母组合成中、英文句子时，往往要受到语法、习惯用语、修辞等的制约，因此中、英文句子中前后出现的汉字、字母往往是有依赖性的。如英文字母 T 后面最常出现 H 和 R，而根本不会出现 Q，F，X。这种依赖性我们称为有记忆。

离散有记忆信源的输出需要用联合概率空间$\{\boldsymbol{X}, q(\boldsymbol{X})\}$来描述，信源输出的消息可表示为符号序列 $\boldsymbol{x} = x_1 x_2 \cdots x_i \cdots$，其中，$x_i$ 表示在 i 时刻信源所发出的符号，i 的数值越小，时间上越早。很明显，有记忆信源在 i 时刻发出的符号与 i 时刻以前信源所发出的符号有关，即由条

件概率 $p(x_i \mid \cdots\ x_{i-2}\ x_{i-1})$确定。

多数有记忆信源的记忆长度是有限的，即某一时刻信源发出的符号只与前面已发出的若干个符号有关。为了描述这种有限的记忆关系，常引入“状态”的概念。这样，信源发出的符号与信源所处状态有关。

下面以马尔可夫信源为例来介绍有记忆信源。

设信源在 r 时刻发出的符号 x_r 与前 m 个符号 x_{r-m}，x_{r-m+1}，…，x_{r-1} 有关（称为 m 阶），这 m 个时间上依次相邻的符号组成一个状态 s，若 $x_i \in \{a_1, a_2, \cdots, a_k\}$，则可能的状态有 k^m 种，即 $s_1, s_2, \cdots, s_{k^m}$。用 e_r 表示 r 时刻符号 x_r 发出前的状态，$e_r = x_{r-m}\ x_{r-m+1} \cdots x_{r-1} = s_i$，当符号 x_r 发出后，状态将发生改变，记为 $e_{r+1} = x_{r-m+1}x_{r-m+2} \cdots x_r = s_j$，用 $p(s_j \mid s_i)$表示 s_i 状态到 s_j 状态的转移概率。

当状态转移概率和已知状态下发出符号的概率与时刻无关，即 $p(e_{r+1} = s_j \mid e_r = s_i) = p(e_{t+1} = s_j \mid e_t = s_i) = p(s_j \mid s_i)$和 $p(x_r = a_l \mid e_r = s_i) = p(x_t = a_l \mid e_t = s_i) = p(a_l \mid s_i)$时，称为时齐的。

若信源输出的序列消息与信源的状态满足下列两个条件，则该信源就称为马尔可夫信源。

（1）某一时刻信源的输出只与当时的信源状态有关，而与以前的状态无关。

有
$$p(x_r = a_l \mid e_r = s_i, e_{r-1} = s_t, e_{r-2} = s_n, \cdots) = p(x_r = a_l \mid e_r = s_i)$$

满足
$$\sum_{l=1}^{k} p(x_r = a_l \mid e_r = s_i) = 1$$

（2）某一时刻信源所处的状态只由前一时刻的输出符号和前一时刻的状态唯一决定。

$$p(e_{r+1} = s_j \mid x_r = a_l, e_r = s_i) = \begin{cases} 0 \\ 1 \end{cases}$$

对于时齐马尔可夫信源，满足

$$\begin{cases} p(s_j) = \sum_i p(s_i)\, p(s_j \mid s_i) \geq 0 \\ \sum_j p(s_j) = 1 \end{cases}$$

【例 1.4】 某二阶时齐马尔可夫信源，设信源符号集为$\{a_1, a_2\}$，状态集为$\{s_1 = a_1\ a_1, s_2 = a_1\ a_2$ 或 $a_2\ a_2, s_3 = a_2\ a_1\}$，各状态之间的转移情况如图 1-2 所示，求各状态的概率分布。

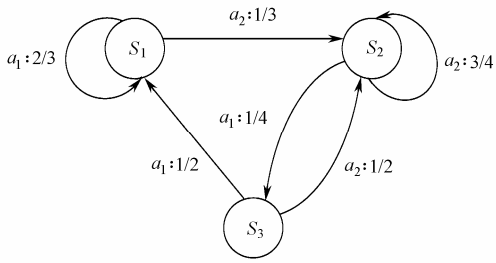


图 1-2　状态转移图

由图 1-2 可知，已知状态下发出符号的概率分别为

$$p(a_1 \mid s_1) = \frac{2}{3} \qquad p(a_2 \mid s_1) = \frac{1}{3}$$

$$p(a_1 \mid s_2)=\frac{1}{4} \qquad p(a_2 \mid s_2)=\frac{3}{4}$$
$$p(a_1 \mid s_3)=\frac{1}{2} \qquad p(a_2 \mid s_3)=\frac{1}{2}$$

状态转移概率分别为

$$p(s_1 \mid s_1)=\frac{2}{3} \qquad p(s_2 \mid s_1)=\frac{1}{3} \qquad p(s_3 \mid s_1)=0$$
$$p(s_1 \mid s_2)=0 \qquad p(s_2 \mid s_2)=\frac{3}{4} \qquad p(s_3 \mid s_2)=\frac{1}{4}$$
$$p(s_1 \mid s_3)=\frac{1}{2} \qquad p(s_2 \mid s_3)=\frac{1}{2} \qquad p(s_3 \mid s_3)=0$$

由于系统是时齐的，由方程组

$$\begin{cases} p(s_1)=\dfrac{2}{3}p(s_1)+\dfrac{1}{2}p(s_3) \\ p(s_2)=\dfrac{1}{3}p(s_1)+\dfrac{3}{4}p(s_2)+\dfrac{1}{2}p(s_3) \\ p(s_3)=\dfrac{1}{4}p(s_2) \\ p(s_1)+p(s_1)+p(s_3)=1 \end{cases}$$

可进一步求出各个状态的分布概率，得 $p(s_1)=\frac{3}{13}$，$p(s_2)=\frac{8}{13}$，$p(s_3)=\frac{2}{13}$。

1.3.3 波形信源

波形信源输出的消息在时间和幅度取值上都是连续的，如语音、图像信号。对于这种信源输出的消息，可用随机过程来描述。常见的波形信源输出的消息是时间上或频率上有限的随机过程，根据取样定理，它可以转换成时间上离散，而每个取样值都连续的随机变量，若对每个取样值再量化处理，就可将连续的取值转换成有限的或可数的离散值。这样波形信源就可转换成离散信源来处理。

连续信源的数学模型是连续型的概率空间，即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} (a,b) \\ q(x) \end{bmatrix}$$

满足

$$\int_a^b q(x)\mathrm{d}x=1$$

式中，$q(x)$为随机变量 x 在取值区间$(a，b)$的概率密度函数。

【例 1.5】 高斯分布信源，其概率统计模型为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} (-\infty,+\infty) \\ \dfrac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{\frac{-(x-m)^2}{2\sigma^2}} \end{bmatrix}$$

1.4 信道及其数学模型

信道是信息传输的通道，如图 1-3 所示，信道有输入端和输出端。信息论中仅关心信道输入、输出之间的关系，而不研究信号在信道中传输的物理过程。因此可以将信道模型看成一个黑匣子。由于干扰的存在，信道的输入和输出之间一般不是确定的函数关系。数字通信系统中只讨论编码译码问题，可以将信道看成一个数字序列的变换器，它将输入消息 x 变换成输出消息 y，以信道转移概率 $p(y|x)$来描述信道的统计特性。

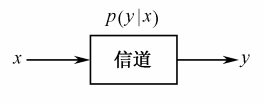


图 1-3　信道模型

信道可以按不同的特性进行分类，根据输入和输出信号的特点可分为以下四类。

- 离散信道：信道的输入和输出都是时间上离散、取值离散的随机序列。离散信道有时也称为数字信道。
- 连续信道：信道的输入和输出都是时间上离散、取值连续的随机序列，又称为模拟信道。
- 半连续信道：输入序列和输出序列中一个是离散的，而另一个是连续的。
- 波形信道：信道的输入和输出都是时间上连续，并且取值也连续的随机信号。

与信源一样，其他信道都可以通过抽样或量化转化为离散信道。

根据统计特性，即转移概率 $p(y|x)$的不同，信道又可分为以下两类。

- 无记忆信道：信道的输出 y 只与当前时刻的输入 x 有关。
- 有记忆信道：信道的输出 y 不仅与当前时刻的输入有关，还与以前的若干个输入及输出消息有关。

实际上，卫星信道和深空信道可近似看成离散无记忆信道。在高频、散射和有线信道中，各种干扰所造成的错误往往不是单个地而是成群成串地出现，也就是一个错误的出现往往引起前后码元的错误，表现为错误之间的相关性，因此它是有记忆信道。

1.4.1 离散无记忆信道

离散无记忆信道（Discrete Memoryless Channel，DMC）的输入和输出消息都是离散无记忆的单个符号，设离散无记忆信道的输入符号 $x_i \in \{a_1, a_2, \cdots, a_k\}$，$1 \leqslant i \leqslant k$，输出符号 $y_j \in \{b_1, b_2, \cdots, b_D\}$，$1 \leqslant j \leqslant D$，信道的特性可表示为转移概率矩阵，即

$$\boldsymbol{P} = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_D|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \cdots & p(y_D|x_2) \\ & \vdots & & \\ p(y_1|x_k) & p(y_2|x_k) & \cdots & p(y_D|x_k) \end{bmatrix}$$

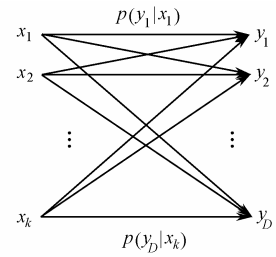


图 1-4　离散无记忆信道

矩阵第 i 行第 j 列的概率 $p(y_j|x_i)$对应为已知输入消息为 x_i，当输出消息为 y_j 时的信道转移概率，满足 $0 \leqslant p(y_j|x_i) \leqslant 1$，且 $\sum_{j=1}^{D} p(y_j|x_i) = 1$。

也可以将信道特性表示成图 1-4 所示的线状图形式。

下面列举几种常见的离散无记忆信道。

1．二元对称信道（Binary Symmetric Channel）

这是一种很重要的信道，它的输入符号 $x \in \{0, 1\}$，输出符号 $y \in \{0, 1\}$，信道特性可表示为信道矩阵，即

$$\boldsymbol{P} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

式中，p 称为信道错误概率。转移概率 $p(y|x)$如图 1-5 所示，即 $p(y=0|x=0) = p(y=1|x=1) = 1-p$，$p(y=1|x=0) = p(y=0|x=1) = p$。

相应有 k 元对称信道，$x, y \in \{0, 1, \cdots, k-1\}$，信道转移概率为

$$p(y|x) = \begin{cases} 1-p & x = y \\ \dfrac{p}{k-1} & x \neq y \end{cases}$$

2．无干扰信道

这是一种最理想的信道，也称为无噪无损信道，信道的输入和输出符号间有确定的一一对应关系，即 $p(y|x) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}$。

在如图 1-6 所示的三元无干扰信道中，x，$y \in \{0, 1, 2\}$，对应信道矩阵是单位矩阵 $\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$。

3．二元删除信道

在实际中，当我们对于接收符号不能做出肯定或否定判决时，可引入删除符号 e，表示对该符号存有疑问，可作为有误或等待得到更多信息时再进行判决。

二元删除信道如图 1-7 所示，输入符号 $x \in \{0, 1\}$，输出符号 $y \in \{0, e, 1\}$，转移概率矩阵为 $\boldsymbol{P} = \begin{bmatrix} p & 1-p & 0 \\ 0 & 1-p & p \end{bmatrix}$。

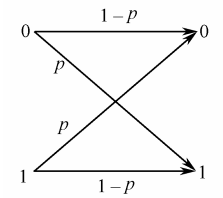


图 1-5　二元对称信道

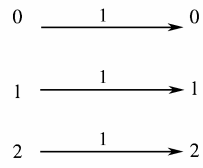


图 1-6　三元无干扰信道

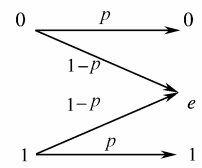


图 1-7　二元删除信道

假如以高、低电平代表输入的 1，0 逻辑，设输出高电平（＞2.4 V）判决为输出 1，输出低电平（＜0.7 V）判决为输出 0。原则上 0.7～2.4 V 输出电平会造成逻辑混乱，是不允许出现的，但由于信道的干扰，输出电平也可能介于 0.7～2.4 V 之间，将这一区间的输出判决为删除符号 e。可以看出，0 错成 1 的可能性要比错成 e 的可能性小得多，同样，1 错成 0 比错成 e

的可能性小得多。应当指出，当对码元进行删除处理时，它在序列中的位置是已知的，仅不知其值是 0 还是 1，故这种信道的纠错要比 BSC 信道容易。

4．二元Z信道

信道输入符号 $x \in \{0, 1\}$，输出符号 $y \in \{0, 1\}$，转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}$ 或 $\boldsymbol{P}=\begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix}$。

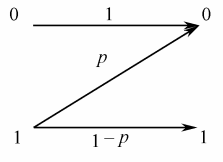


图 1-8　二元 Z 信道

1.4.2　离散无记忆扩展信道

N 维离散扩展信道的输入和输出都是长为 N 的符号序列消息，如图 1-9 所示，$\boldsymbol{x}=x_1\,x_2\cdots x_N$，$\boldsymbol{y}=y_1\,y_2\cdots y_N$。

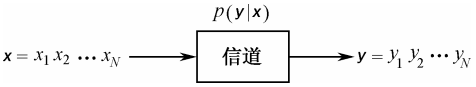


图 1-9　N 维扩展信道

若 $x_i \in \{a_1, a_2, \cdots, a_k\}$，$y_j \in \{b_1, b_2, \cdots, b_D\}$，$1 \leqslant i, j \leqslant N$，则长为 N 的输入序列消息集为 $\boldsymbol{X}=\{a_1a_1\cdots a_1, a_1a_1\cdots a_2, \cdots, a_ka_k\cdots a_k\}$，$\boldsymbol{x} \in \boldsymbol{X}$，输出序列消息集为 $\boldsymbol{Y}=\{b_1b_1\cdots b_1, b_1b_1\cdots b_2, \cdots, b_Db_D\cdots b_D\}$，$\boldsymbol{y} \in \boldsymbol{Y}$。信道的特性用序列的转移概率 $p(\boldsymbol{y}|\boldsymbol{x})=p(y_1\,y_2\cdots y_N|\,x_1\,x_2\cdots x_N)$ 描述。

当信道无记忆时，$p(\boldsymbol{y}|\boldsymbol{x})=\prod_{i=1}^{N} p(y_i|x_i)$，满足 $0 \leqslant p(\boldsymbol{y}|\boldsymbol{x}) \leqslant 1$，且 $\sum_{\boldsymbol{Y}} p(\boldsymbol{y}|\boldsymbol{x})=1$。

【例 1.6】　求二元对称信道的二维扩展信道。

解：二元对称信道的输入符号 $x \in \{0, 1\}$，输出符号 $y \in \{0, 1\}$，转移概率 $p(0|0)=p(1|1)=1-p$，$p(1|0)=p(0|1)=p$，二维扩展后信道的输入和输出消息都是长为 2 的符号序列，$\boldsymbol{x} \in \{00, 01, 10, 11\}$，$\boldsymbol{y} \in \{00, 01, 10, 11\}$。

可以计算出序列的转移概率 $p(\boldsymbol{y}|\boldsymbol{x})$，分别为

$$p(y_1\,y_2=00|\,x_1\,x_2=00)=p(y_1=0|\,x_1=0)\cdot p(y_2=0|\,x_2=0)=(1-p)^2$$

$$p(01|00)=p(0|0)\cdot p(1|0)=(1-p)p$$

$$\vdots$$

$$p(10|01)=p(1|0)\cdot p(0|1)=p^2$$

$$\vdots$$

$$p(11|11)=p(1|1)\cdot p(1|1)=(1-p)^2$$

表示成矩阵为　$\boldsymbol{P}=\begin{bmatrix} (1-p)^2 & (1-p)p & p(1-p) & p^2 \\ (1-p)p & (1-p)^2 & p^2 & p(1-p) \\ p(1-p) & p^2 & (1-p)^2 & (1-p)p \\ p^2 & p(1-p) & (1-p)p & (1-p)^2 \end{bmatrix}$

本 章 小 结

本章是信息论的基本概念，介绍的主要内容有：

（1）信息是关于事物运动的状态和规律。消息是能被人们感觉器官感知的客观物质和主观思维的运动状态或存在状态，消息中包含信息。从通信的角度讲，信息论是应用近代概率统计方法研究狭义信息的度量方法，研究各种信源、信道的描述和信源、信道的编码定理。

（2）数字通信系统由信源、信源及信道编码器、信道、信源及信道译码器、信宿五部分组成。信源编码器将信源的冗余度剔除以提高通信的有效性，信道编码器通过增加冗余的纠错、检错码元来提高通信的可靠性。

（3）信源分为离散的和连续的、无记忆的和有记忆的。信源的数学模型为一个样本空间及其概率测度$\{\boldsymbol{X}, q(\boldsymbol{X})\}$。离散无记忆信源每次只输出一个离散的符号消息，不同时刻发出的符号之间彼此统计独立；离散无记忆扩展信源输出的消息是由许多不同时刻发出的符号所组成的符号序列，先后发出的符号彼此间统计独立，符号序列的概率是序列中各个符号概率的乘积；有记忆信源时间上先后发出的消息间有依赖关系。

（4）信道分为离散的和连续的、无记忆的和有记忆的。信道的数学模型用转移概率描述。离散无记忆信道的输入和输出都是离散无记忆的单个符号。离散无记忆 N 维扩展信道的输入和输出消息都是长为 N 的符号序列，序列的信道转移概率是对应 N 个符号的信道转移概率的乘积。

思考题与习题

1.1　信息与消息的概念有何区别？

1.2　信息论研究的范畴是什么？

1.3　简述数字通信系统五个组成部分的作用。

1.4　数字通信系统的主要性能指标有哪些？常采用哪些措施来提高这些性能？

1.5　什么是误码率？什么是误信率？两者有何关系？

1.6　什么是码元速率？什么是信息速率？两者有何关系？

1.7　同时掷一对均匀的骰子，要得知面朝上点数之和，描述这一信源的数学模型。

1.8　设有一个二阶二元马尔可夫信源，其 4 种状态的转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \end{bmatrix}$$

画出状态图并求稳态下各状态出现的概率。

1.9　有一个二元对称信道，信道误码率 $p=0.06$，设该信道以 1000 个符号/秒的速率传输输入符号，现有一消息序列，共有 9500 个符号，并设消息中 $q(0)=q(1)=0.5$，问从信息传输的角度来考虑，10 秒能否将消息无失真地传送完？

1.10　求二元删除信道的二次扩展信道。

第 2 章
信息的度量

内容提要

根据香农对于信息的定义，信息是对系统不确定性的度量，尤其在通信系统中，研究的是信息的处理、传输和存储，所以对于信息的定量计算是非常重要的。本章主要从通信系统模型入手，研究在离散情况下各种信息的描述方法及定量计算，讨论它们的性质和相互关系。

知识要点

熵、平均互信息量、极大离散熵、平均互信息量与信源及信道的关系。

教学建议

本章涉及的名词、物理量比较多，理顺这些量之间的关系，通过例题搞清楚它们的物理意义，是进行以后各章学习的基础。希望对文中的定理 2.1 及定理 2.2 重点关注，因为它们是香农第一、第三定理的理论依据，而香农的三大定理奠定了整个信息论的基础。另外本章用到概率统计的一些知识，需要预先复习。建议教学时数为 10 学时。

2.1 自信息量和互信息量

这里引入随机事件不确定性的概念，由于各种事件出现的概率不同，它们所包含的不确定性也有大小的差别，而一个事件的自信息量就是对其不确定性的度量。互信息量则表明了两个随机事件的相互约束程度。

对于随机事件集 $X=\{x_1, x_2, \cdots, x_i, \cdots, x_I\}$ 中的随机事件 x_i，其出现概率记为 $q(x_i)$，由概率定义，$q(x_i)$ 应满足

$$\begin{cases} q(x_i) \geq 0 & i=1,2,\cdots,I \\ \sum_{i=1}^{I} q(x_i)=1 \end{cases}$$

同样，对于每一随机变量 $y_j \in Y, j=1, 2, \cdots, J$，其出现概率记为 $\omega(y_j)$，由概率定义，$\omega(y_j)$ 应满足

$$\begin{cases} \omega(y_j) \geq 0 & j=1,2,\cdots,J \\ \sum_{j=1}^{J} \omega(y_j)=1 \end{cases}$$

将两个事件 x_i, y_j 同时出现的概率记为 $p(x_i y_j)$，则 $p(x_i y_j)$ 应满足

$$\begin{cases} p(x_i y_j) \geq 0 \\ \sum_{i=1}^{I}\sum_{j=1}^{J} p(x_i y_j)=1 \end{cases}$$

相应的条件概率为

$$\begin{cases} \phi(x_i | y_j)=\dfrac{p(x_i y_j)}{\omega(y_j)} \\ p(y_j | x_i)=\dfrac{p(x_i y_j)}{q(x_i)} \end{cases}$$

满足

$$\begin{cases} \sum_{i=1}^{I} \phi(x_i | y_j)=1 \\ \sum_{j=1}^{J} p(y_j | x_i)=1 \end{cases}$$

下面的关系式成立

$$\begin{cases} \sum_{i=1}^{I} p(x_i y_j)=\omega(y_j) \\ \sum_{j=1}^{J} p(x_i y_j)=q(x_i) \end{cases}$$

【例 2.1】 对一人群进行性别和年龄的统计，分别用 x_i 代表性别、用 y_j 代表年龄，随机抽取一人，其性别为 x_i 的概率为 $q(x_i)$，年龄为 y_j 的概率为 $\omega(y_j)$，在人群中任取一人，其性别

为 x_i 同时年龄又为 y_j 的概率为 $p(x_iy_j)$，$p(y_j \mid x_i)$则表示已知一个人性别为 x_i 而年龄恰好为 y_j 的概率，$\phi(x_i \mid y_j)$则表示已知一个人年龄为 y_j 而性别恰好为 x_i 的概率。

2.1.1 自信息量和条件自信息量

随机信息源含有不确定性，若通过某个过程对信息源有一定的了解，也就是从信息源获得了信息，那么从信息源获得信息的过程就是不确定性减少的过程。从通信的角度来说，通信是为了将某个消息从信源传到信宿，这个消息中一定包含着信宿未知的一些信息（即存在一定的不确定性），经过通信，信宿得到了一定量的信息，所以通信是不确定性减少的过程。那么传递一次得到多少信息呢，直观地可把信息量定义为

收到某消息获得的信息量 = 不确定性减少的量

而事件 x 发生的不确定性与事件发生的概率 $q(x)$有关，概率越小，不确定性就越大。因此将某事件发生所得到的信息量记为 $I(x)$，$I(x)$应该是该事件发生的概率的函数，即

$$I(x) = f[q(x)]$$

1．自信息量

直观地看，自信息量的定义应满足以下四点：

① $I(x)$应该是 $q(x)$的单调递减函数，概率小的事件一旦发生则赋予的信息量大，概率大的事件如果发生则赋予的信息量小；

② 信息量应具有可加性，对于两个独立事件，其信息量应等于各事件自信息量之和；

③ 当 $q(x) = 1$ 时，$I(x) = 0$，表示确定事件发生得不到任何信息；

④ 当 $q(x) = 0$ 时，$I(x) \to \infty$，表示不可能事件一旦发生，信息量将无穷大。

综合上述条件，数学上可证明，对数函数可以满足上述要求，将自信息量定义为

$$I(x) \triangleq -\log q(x) \tag{2-1}$$

这里要说明的是，上面的定义是合理的，因为随机事件是否发生蕴涵着不确定性，一个随机事件出现概率越接近 1，该事件发生的可能性越大，它所包含的不确定性就越小，事件发生给收信者提供的信息量就越小。反之，一个随机事件的出现概率越接近 0，该事件发生的可能性越小，它所包含的不确定性就越大，事件发生给收信者提供的信息量就越大。

例如，2003 年国家规定“五一”放七天长假，如果老师对学生宣布“5 月 1 日至 5 月 7 日这七天不上课。”，因为是已知事件，学生并不奇怪，他们没有获得多少信息；但 2003 年 SARS 狂虐期间，为了防止人口流动引发 SARS 大面积传播，国家决定大专院校不放假，老师对学生宣布“根据上级精神，五一期间不放假，仍旧上课”，因为这是偶然事件，发生概率极小，故学生将从中获得信息。

【例 2.2】 在一个盒子中放有 n 个阻值各为 1 Ω，2 Ω，…，n Ω的电阻，将从盒子中取出阻值为 i Ω的电阻记为事件 x_i（i =1，2，…，n），则 x_i 发生的概率为 $q(x_i) = \dfrac{1}{n}$（等概分布），这一事件发生所获得的信息量为 $I(x_i) = -\log\dfrac{1}{n} = \log n$。

【例 2.3】 若盒中有 6 个电阻，阻值为 1 Ω、2 Ω、3 Ω的电阻分别为 2 个、1 个、3 个，将从盒子中取出阻值为 i Ω的电阻记为事件 x_i（i = 1，2，3），组成事件集 $X = \{x_1, x_2, x_3\}$，其

概率分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$，计算出各事件的自信息量，如表 2-1 所示。

表 2-1　事件概率分布及自信息量

消息 x_i	x_1	x_2	x_3
概率分布 $q(x_i)$	1/3	1/6	1/2
自信息量 $I(x_i)$	log 3	log 6	log 2

自信息量 $I(x_i)$代表两种含义：

① 事件 x_i 发生以前，表示事件发生的先验不确定性，一个事件不常出现，它的概率就小，当该事件发生时收信者获得的信息就多，或者说事件所携带的自信息量大，因此也可以说自信息量是随机事件的一个固有特征；

② 当事件 x_i 发生以后，表示事件 x_i 所能提供的最大信息量（在无噪情况下）。

例如，$I(x_1) = \log 3$ 就是能否取出 1 Ω电阻的不确定性的度量。如前所述，事件发生的过程就是不确定性减少的过程，如果取出一个电阻后，获得了 log3 的信息，那就把所有不确定性都排除了，就能唯一地确定取出的电阻就是阻值为 1 Ω的电阻。换言之，$I(x_i)$是唯一地确定事件 x_i 所必须提供的信息量。

从表 2-1 可以看出，概率小的事件携带的自信息量大，下面的例子可以更明确地说明这一点。

【例 2.4】　$X = \{0, 1\}$为随机变量，其概率分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.01 & 0.99 \end{bmatrix}$，计算出各消息的自信息量，如表 2-2 所示。

表 2-2　消息概率分布及自信息量

消息 x_i	0	1
概率分布 $q(x_i)$	0.01	0.99
自信息量 $I(x_i)$	6.644（比特）	0.014（比特）

从表中可看出，由于事件 $x = 0$ 出现的概率比事件 $x = 1$ 出现的概率小得多，故对应的自信息量要大得多。

显然，由于概率 $0 \leqslant q(x_i) \leqslant 1$，则

$$I(x_i) = -\log q(x_i) \geqslant 0 \qquad (2\text{-}2)$$

自信息量的单位与 log 函数所选用的对数底数有关，如底数分别取 2，e，10，则自信息量单位分别为比特（bit, binary unit 的缩写）、奈特（nat, natural unit 的缩写）和哈特（Hart, Hartley 的缩写，以纪念哈特莱首先提出用对数来度量信息）。

三者之间的互换关系如下：

$$1 \text{ 奈特} = \log_2 \mathrm{e} \text{ 比特} = 1.443 \text{ 比特}$$

$$1 \text{ 哈特} = \log_2 10 \text{ 比特} = 3.322 \text{ 比特}$$

比特是信息论中常用到的单位，在本书中若不另加说明，则默认取 2 为对数底数，信息量的单位为比特（bit），为了简洁常把底数 2 略去不写。

【例 2.5】　信源消息 $X=\{0, 1\}$，信源等概分布，计算出自信息量，如表 2-3 所示。

表 2-3　两个等可能事件的自信息量

x_i	0	1
$Q(x_i)$	1/2	1/2
$I(x_i)$	log2	log2

可以看出，$\log 2=1$（比特），1 比特信息量就是两个互不相容的等可能事件之一发生时所提供的信息量。

$I(x_i) = -\log q(x_i)$是针对一维空间的，可把它推广到二维空间$\{XY, p(x_i y_j)\}$，式中，$p(x_i y_j)$为元素 $x_i\ y_j$ 在二维空间 XY 上的联合概率密度，在二维联合集 XY 上元素 $x_i\ y_j$ 的联合自信息量 $I(x_i y_j)$定义为

$$I(x_i y_j) \triangleq -\log p(x_i y_j) \tag{2-3}$$

2．条件自信息量

在已知事件 y_j 条件下，随机事件 x_i 发生的概率为条件概率 $\phi(x_i \mid y_j)$，条件自信息量 $I(x_i \mid y_j)$ 定义为

$$I(x_i \mid y_j) \triangleq -\log \phi(x_i \mid y_j) \tag{2-4}$$

与事件 x_i 的自信息量类似，随机事件的条件自信息量也可以理解为，这是在事件 y_j 给定的条件下关于事件 x_i 是否发生的平均不确定性，若条件概率 $\phi(x_i \mid y_j)$小，则给定 y_j 时关于 x_i 的是否发生有着较大的不确定性，反之不确定性就小。同样，$I(x_i \mid y_j)$也可以看成在事件 y_j 给定的条件下，唯一地确定事件 x_i 所必须提供的信息量。

显然也有

$$I(x_i \mid y_j) \geqslant 0 \tag{2-5}$$

条件自信息量的单位也由 log 函数所选用的对数底数决定，当底数分别为 2，e, 10 时，条件自信息量的单位分别为比特（bit）、奈特（nat）、哈特（Hart）。

【例 2.6】　某住宅区共建有若干栋商品房，每栋有 5 个单元，每个单元住有 12 户，甲要到该住宅区找朋友乙，若

（1）甲只知道乙住在第 5 栋，他找到乙的概率有多大？他能得到多少信息？

（2）甲除知道乙住在第 5 栋外，还知道乙住在第 3 单元，他找到乙的概率又有多大？他能得到多少信息？

解：因每栋有 5 个单元，每个单元又有 12 户，除此之外甲并不知道其他信息，因此对甲来说第 5 栋共有 60 户是等概分布的，用 x_i 代表单元数，y_j 代表户号。

（1）甲找到乙这一事件是二维联合集 XY 上的等概分布 $p(x_i y_j)=\dfrac{1}{60}$，这一事件提供给甲的信息量为

$$I(x_i y_j) = -\log p(x_i y_j) = \log 60 = 5.907\text{（比特）}$$

（2）在二维联合集 XY 上的条件分布概率为 $p(y_j \mid x_i)=\dfrac{1}{12}$，这一事件提供给甲的信息量为条件自信息量

$$I(y_j \mid x_i) = -\log p(y_j \mid x_i) = \log 12 = 3.585\text{（比特）}$$

2.1.2 互信息量和条件互信息量

1．互信息量

下面从通信的角度引出互信息量的概念，一次通信过程可以简单地用图 2-1 所示的模型来表示。信源包含若干个消息，分别用信源符号 $X = \{x_1, x_2, \cdots, x_I\}$ 表示，$x_i \in \{a_1, a_2, \cdots, a_k\}$，$i = 1, 2, \cdots, I$，由于事先并不知道信源会发出哪个消息，所以每个信源符号 x_i 相当于一个随机事件。就图 2-1 所示的简单通信系统而言，也常把信源的输出符号称为信道的输入符号，而把信宿接收到的符号称为信道的输出符号，经过信道传输，信宿方接收到符号 $Y = \{y_1, y_2, \cdots, y_J\}$，$y_J \in \{b_1, b_2, \cdots, b_D\}$，$j = 1, 2, \cdots, J$。

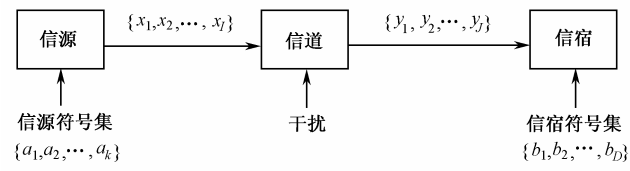


图 2-1　简单的通信模型

消息 x_i 的概率分布 $q(x_i)$ 称为先验概率，接收到符号 y_j 后，接收者重新估计事件 x_i 发生的概率，记为条件概率 $\phi(x_i \mid y_j)$，也称 $\phi(x_i \mid y_j)$ 为后验概率。

事件 x_i 是否发生具有不确定性，用 $I(x_i)$ 度量。接收到符号 y_j 后，事件 x_i 是否发生仍保留有一定的不确定性，用 $I(x_i \mid y_j)$ 度量。观察事件前后，这两者之差就是通信过程中所获得的信息量，用 $I(x_i; y_j)$ 表示，即

$$I(x_i; y_j) = I(x_i) - I(x_i \mid y_j) = \log \frac{\phi(x_i \mid y_j)}{q(x_i)} \tag{2-6}$$

称式（2-6）为事件 x_i 和事件 y_j 之间的互信息量。

根据概率互换公式 $p(x_i y_j) = p(y_j \mid x_i) q(x_i) = \phi(x_i \mid y_j) \omega(y_j)$（在图 2.1 所示的简单通信模型中，$\omega(y_j)$ 表示信宿收到 y_j 的概率；$p(y_j \mid x_i)$ 是信源发出 x_i 符号，而信宿收到 y_j 的概率，即信道转移概率），互信息量 $I(x_i; y_j)$ 有多种表达形式，即

$$I(x_i; y_j) = \log \frac{p(y_j \mid x_i)}{\omega(y_j)} = I(y_j) - I(y_j \mid x_i) \tag{2-7}$$

$$I(x_i; y_j) = \log \frac{p(x_i y_j)}{q(x_i)\omega(y_j)} = I(x_i) + I(y_j) - I(x_i y_j) \tag{2-8}$$

仍以图 2-1 所示简单通信模型为例，从式（2-7）看 $I(x_i; y_j)$ 的物理意义。这相当于观察者站在信源方来观察问题，在信源发出信号前，信宿收到 y_j 的概率为 $\omega(y_j)$，其不确定性用 $I(y_j)$ 度量。而信源发出符号 x_i 后，由于干扰，“信宿收到 $Y = \{y_1, y_2, \cdots, y_J\}$ 中的哪个符号”这一事件具有发散性，即信宿是否收到 y_j 仍存有不确定性，用 $I(y_j \mid x_i)$ 度量。这二者之差就是事件发生过程中观察者所获得的信息量 $I(x_i; y_j) = I(y_j) - I(y_j \mid x_i)$。

从式（2-8）看 $I(x_i; y_j)$ 的物理意义，通信前 $X = \{x_1, x_2, \cdots, x_I\}$ 和 $Y = \{y_1, y_2, \cdots, y_J\}$ 统计独立，联合概率为 $p(x_i y_j) = q(x_i)\omega(y_j)$，不确定性用 $-\log q(x_i)\omega(y_j) = I(x_i) + I(y_j)$ 度量。通

信后由于信道转移概率 $p(y_j \mid x_i)$ 的存在，符号 x_i y_j 有了某种关联，联合概率 $p(x_i y_j) = p(y_j \mid x_i) q(x_i)$，发 x_i 收 y_j 的不确定性用 $I(x_i\ y_j) = -\log p(x_i\ y_j)$ 度量，二者之差就是通信过程中，x_i 与 y_j 所得到的互信息量。

注意：式（2-6）的 $I(x_i\ ; y_j)$ 和式（2-3）的 $I(x_i\ y_j)$ 的区别在于，前者是事件 $x_i \in X$ 和事件 $y_j \in Y$ 之间的互信息量，后者是二维空间 XY 上元素 $x_i y_j$ 的自信息量。

同样，互信息量的单位也由 log 函数所取的对数底数确定，当底数分别为 2，e，10 时，条件自信息量的单位分别为比特（bit）、奈特（nat）、哈特（Hart）。

事件 x_i 和 y_j 之间之所以有互信息，是因为两个事件之间统计相关，当事件 x_i 和 y_j 相互统计独立时，有 $\begin{cases} \phi(x_i \mid y_j) = q(x_i) \\ p(y_j \mid x_i) = \omega(y_j) \\ p(x_i y_j) = q(x_i)\omega(y_j) \end{cases}$，则根据式（2-6）或式（2-7）或式（2-8），都可得到 $I(x_i ; y_j) = 0$，这说明当两事件独立时，不能从对一个事件的观察获得另一事件的任何信息。

将事件互信息量的概念推广至多维空间。

在三维 XYZ 联合集中，有

$$\begin{aligned} I(x_i ; y_j z_k) &= \log \frac{p(x_i \mid y_j z_k)}{q(x_i)} \\ &= \log \frac{p(x_i \mid y_j z_k)}{p(x_i / y_j)} \cdot \frac{p(x_i / y_j)}{q(x_i)} \\ &= \log \frac{p(x_i \mid y_j)}{q(x_i)} + \log \frac{p(x_i \mid y_j z_k)}{p(x_i \mid y_j)} \\ &= I(x_i ; y_j) + I(x_i ; z_k \mid y_j) \end{aligned}$$

即

$$I(x_i ; y_j\ z_k) = I(x_i ; y_j) + I(x_i ; z_k \mid y_j) \qquad (2\text{-}9)$$

即一对事件 y_j z_k 发生后，与事件 x_i 之间的互信息量，等于事件 y_j 与 x_i 之间的互信息量，加上在事件 y_j 已知的条件下，事件 z_k 与 x_i 之间的互信息量。

类似地，在 N 维 $U_1 U_2 \cdots U_N$ 联合空间，有

$$\begin{aligned} I(u_1 ; u_2 u_3 \cdots u_N) = I(u_1 ; u_2) + I(u_1 ; u_3 \mid u_2) + \cdots + I(u_1 ; u_i \mid u_2 \cdots u_{i-1}) + \cdots + \\ I(u_1 ; u_N \mid u_2 \cdots u_{N-1}) \end{aligned} \qquad (2\text{-}10)$$

【例 2.7】 有 A, B, C, D 四人，A 约他的三个朋友 B，C，D 当晚一起到茶座闲谈，三人都答应无特殊事件一定来，视为等概事件 $q(\mathrm{B}) = q(\mathrm{C}) = q(D) = \frac{1}{3}$。上午 A 接到 B 的电话，因事不能来，记为事件 x_1，条件概率 $p(B \mid x_1) = 0$，$p(C \mid x_1) = p(D \mid x_1) = \frac{1}{2}$；下午 A 又接到 C 的电话，因事不能来，记为事件 x_2，条件概率 $p(B \mid x_1 x_2) = p(C \mid x_1 x_2) = 0$，$p(D \mid x_1 x_2) = 1$。下面计算 A 获得关于 B, C, D 的各种互信息量。

接到 B 的电话后，A 获得关于 B 的互信息量为

$$I(B ; x_1) = \log \frac{p(B \mid x_1)}{q(B)} = \log \frac{0}{1/3} \to \infty$$

通俗地理解为，B 已明确表示不来，一旦来了，A 大吃一惊，所获信息量为无穷。

接到 B 的电话后，A 获得关于 C, D 的互信息量为

$$I(C;x_1)=\log\frac{p(C\mid x_1)}{q(C)}=\log\frac{1/2}{1/3}=\log 1.5=0.585 \text{（比特）}$$

$$I(D;x_1)=\log\frac{p(D\mid x_1)}{q(D)}=\log\frac{1/2}{1/3}=\log 1.5=0.585 \text{（比特）}$$

接到两个电话后，A 获得关于 B, C, D 的互信息量为

$$I(B;x_1x_2)=\log\frac{p(B\mid x_1x_2)}{q(B)}=\log\frac{0}{1/3}\to\infty$$

$$I(C;x_1x_2)=\log\frac{p(C\mid x_1x_2)}{q(C)}=\log\frac{0}{1/3}\to\infty$$

$$I(D;x_1x_2)=\log\frac{p(D\mid x_1x_2)}{q(D)}=\log\frac{1}{1/3}=\log 3=1.585 \text{（比特）}$$

2．条件互信息量

三维 XYZ 联合集中，在给定条件 z_k 的情况下，若事件 x_i 的后验概率为 $p(x_i\mid z_k)$，并且给定条件 y_j, z_k 的情况下，事件 x_i 的后验概率为 $p(x_i\mid y_jz_k)$，则给定条件 z_k 后，x_i，y_j 的互信息量 $I(x_i;y_j\mid z_k)$定义为

$$I(x_i;y_j\mid z_k)=\log\frac{p(x_i\mid y_jz_k)}{p(x_i\mid z_k)} \qquad (2\text{-}11)$$

3．互信息量的性质

（1）互易性

$$I(x_i;y_j)=I(y_j;x_i) \qquad (2\text{-}12)$$

这从式（2-8）的对称性可得出。

（2）可加性

根据式（2-10）有

$$I(u_1;u_2u_3\cdots u_N)=I(u_1;u_2)+I(u_1;u_3\mid u_2)+\cdots+I(u_1;u_i\mid u_2\cdots u_{i-1})+\cdots+I(u_1;u_N\mid u_2\cdots u_{N-1})$$

信息的可加性对于计算多个事件之间的互信息量非常方便，特别是在多用户信息论中十分有用。

（3）当 x_i ,y_j 统计独立时，互信息量 $I(x_i;y_j)=0$ 及条件互信息量 $I(x_i;y_j\mid z_k)=0$

证明：当 x_i , y_j 统计独立时，有 $p(x_i\ \ y_j)=q(x_i)\,\omega(y_j)$ 及 $\phi(x_i\mid y_j)=q(x_i)$。

则
$$I(x_i;y_j)=\log\frac{p(x_iy_j)}{q(x_i)\omega(y_j)}=\log\frac{q(x_i)\omega(y_j)}{q(x_i)\omega(y_j)}=\log 1=0$$

同样也有
$$I(x_i;y_j\mid z_k)=\log\frac{p(x_i\mid y_jz_k)}{p(x_i\mid z_k)}=\log\frac{p(x_i\mid z_k)}{p(x_i\mid z_k)}=\log 1=0$$

（4）互信息量 $I(x_i;y_j)$可以是正数，也可以是负数

由式（2-6）知，互信息量 $I(x_i;y_j)=\log\frac{\phi(x_i\mid y_j)}{q(x_i)}$，所以 $I(x_i;y_j)$的正、负视比值 $\frac{\phi(x_i\mid y_j)}{q(x_i)}$ 而

定，若事件 y_j 的出现有助于肯定事件 x_i 的出现，即 $\phi(x_i \mid y_j) > q(x_i)$，则比值大于 1，互信息量取正值。反之，若事件 y_j 的出现告知的是 x_i 出现的可能性更小了，即 $\phi(x_i \mid y_j) < q(x_i)$，则比值小于 1，互信息量取负值。从通信角度来看，视 x_i 为发送符号，y_j 为接收符号，$\phi(x_i \mid y_j) < q(x_i)$，说明收到 y_j 后使发送是否为 x_i 的不确定性更大，这是由信道干扰所引起的。

但平均互信息量 $I(X;Y) = \sum_i \sum_j p(x_i y_j) I(x_i; y_j) \geq 0$，后面会讲到。

（5）两个事件的互信息量不大于单个事件的自信息量，即有

$$I(x_i; y_j) \begin{cases} \leq I(x_i) \\ \leq I(y_j) \end{cases} \tag{2-13}$$

证明：

$$I(x_i; y_j) = \begin{cases} \log \dfrac{\phi(x_i \mid y_j)}{q(x_i)} \leq \log \dfrac{1}{q(x_i)} = I(x_i) \\ \log \dfrac{p(y_j \mid x_i)}{\omega(y_j)} \leq \log \dfrac{1}{\omega(y_j)} = I(y_j) \end{cases}$$

类似地也有

$$I(x_i; y_j \mid z_k) \begin{cases} \leq I(x_i \mid z_k) \\ \leq I(y_j \mid z_k) \end{cases} \tag{2-14}$$

式（2-13）说明自信息量 $I(x_i)$（或 $I(y_j)$）是为了确定事件 x_i（或 y_j）的发生所必须提供的信息量，它也是其他任何事件所能提供的关于事件 x_i（或 y_j）的最大信息量，即任何两个事件之间的互信息量不可能大于其中任一事件的自信息量。

下面举例说明在通信过程当中，信源发出的消息与信宿方接收到的符号之间的后验概率及互信息量是如何逐步发生变化的。

【例 2.8】 信源包含 8 个消息 $\{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$，信源编码器将其对应编成 8 个 3 位二进制数 000，001, …, 111。各消息的先验概率已知，在接收过程中，每收到一个数字，各消息的后验概率都相应地发生变化。考虑在接收 100 三个数字的过程中，各后验概率的变化，计算信息量 $I(x_4;100)$。表 2-4 为 8 个 3 位二进制数对应的各种概率。

表 2-4　消息的先验概率和接收 100 过程中各消息的后验概率

信源消息	码　字	消息先验概率	消息后验概率		
			收到 1 后	收到 10 后	收到 100 后
x_0	000	1/16	0	0	0
x_1	001	1/16	0	0	0
x_2	010	1/16	0	0	0
x_3	011	1/16	0	0	0
x_4	100	1/8	1/6	1/2	1
x_5	101	1/8	1/6	1/2	0
x_6	110	1/4	1/3	0	0
x_7	111	1/4	1/3	0	0

根据给定的先验概率，可算出

$$p(x_4)=\frac{1}{8}$$

$$p(x_4|1)=\frac{1/8}{1/8+1/8+1/4+1/4}=\frac{1}{6}$$

$$p(x_4|10)=\frac{1/6}{1/6+1/6}=\frac{1}{2}$$

$$p(x_4 \mid 100)=1$$

将各种后验概率的计算结果列于表 2-4 中，再根据式（2-10）计算出互信息量

$$\begin{aligned}I(x_4;100)&=I(x_4;1)+I(x_4;0 \mid 1)+I(x_4;0 \mid 10)\\&=\log\frac{p(x_4|1)}{p(x_4)}+\log\frac{p(x_4|10)}{p(x_4|1)}+\log\frac{p(x_4|100)}{p(x_4|10)}\\&=\log\frac{1/6}{1/8}+\log\frac{1/2}{1/6}+\log\frac{1}{1/2}\\&=\log 8=3\ \text{（比特）}\end{aligned}$$

也可直接计算出 $I(x_4;100)=\log\dfrac{p(x_4|100)}{p(x_4)}=\log\dfrac{1}{1/8}=3$（比特）

2.2 离散集的平均自信息量

2.2.1 信息熵

1．平均自信息量（熵）

前面讨论的自信息量是针对个别事件的，得到的结论是，要唯一地确定事件 $x_i(i=1,2,\cdots)$，所需要的自信息量为 $I(x_i)=-\log q(x_i)$。实际信源往往包含着许多消息 $X=(x_1, x_2, \cdots)$，而人们注意的也往往是整个系统的统计特性，所以应该考虑整个信源自信息量的统计平均值。当信源各个消息的出现概率相互统计独立时，这种信源称为无记忆信源，无记忆信源的平均自信息量定义为各消息自信息量的概率加权平均值（统计平均值），即平均自信息量 $H(X)$定义为

$$H(X)\triangleq\sum_i q(x_i)I(x_i)=-\sum_i q(x_i)\log q(x_i) \qquad (2\text{-}15)$$

因为 $I(x_i)$是唯一确定事件 x_i所需要的信息量，则 $H(X)$就是唯一确定集合 X 中任一事件 x_i（$i=1$，2，…）所需要的平均信息量，它反映了 X 中事件 x_i出现的平均不确定性。

该平均信息量 $H(X)$的表达式（2-15）与统计物理学中的热熵具有类似的形式，热熵是一个物理系统杂乱性（无序性）的度量。在概念上二者也具有相同之处，故借用熵这个词把 $H(X)$称为集合 X 的信息熵，简称熵。由式（2-15）知，熵 $H(X)$是信源消息概率分布 $q(x_i)$（$i=1$，2，…）的函数。

在例 2.8 中，可算得

$$H(X)=-4\times\frac{1}{16}\log\frac{1}{16}-2\times\frac{1}{8}\log\frac{1}{8}-2\times\frac{1}{2}\log\frac{1}{2}=2.75\ \text{（比特/符号）}$$

【例 2.9】　计算下列信源的熵。

（1）信源一：$\begin{bmatrix} X_1 \\ q(X_1) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 \\ 0.99 & 0.01 \end{bmatrix}$

熵　　$H(X_1) = -0.99\log 0.99 - 0.01\log 0.01 = 0.08$（比特/符号）

（2）信源二：等概信源 $\begin{bmatrix} X_2 \\ q(X_2) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 \\ 0.5 & 0.5 \end{bmatrix}$

熵　　$H(X_2) = -0.5\log 0.5 - 0.5\log 0.5 = 1$（比特/符号）

（3）信源三: 等概信源 $\begin{bmatrix} X_3 \\ q(X_3) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$

熵　　$H(X_3) = -4\times 0.25\log 0.25 = \log 4 = 2$（比特/符号）

（4）信源四：信源为确定事件 $\begin{bmatrix} X_4 \\ q(X_4) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 \\ 0 & 1 \end{bmatrix}$

熵　　$H(X_4) = -0\log 0 - 1\log 1 = 0$

计算结果说明确定事件的熵为零。

（5）信源五：在一般情况下，二元信源的概率分布为 $\begin{bmatrix} X_5 \\ q(X_5) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ \delta & 1-\delta \end{bmatrix}$

熵　　$H(X) = -\delta\log\delta-(1-\delta)\log(1-\delta)$

记　　$H_2(\delta) = -\delta\log\delta-(1-\delta)\log(1-\delta)$

$H_2(\delta)$与δ的关系如图 2-2 所示。

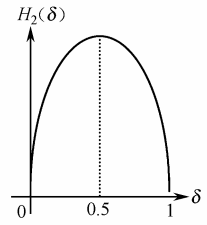


图 2-2　$H_2(\delta)$ 与δ 的关系

比较信源一和信源二的熵可以看出，尽管两个信源含有相同的消息数目，但第一个信源的消息概率分布相差很大，接近确定事件，而信源熵反映的是一种平均不确定程度，因此它的信源熵小；第二个信源的输出是两个互不相容的等可能事件，事件的概率分布相等，事件发生的平均不确定程度很大，因此它的信源熵大。

比较信源二和信源三的熵可以看出，尽管两个信源都为等概分布，但第二个信源只含 2 个事件，而第三个信源含 4 个事件，平均不确定性比第二个信源大，而信源熵反映的是一种平均不确定程度，因此第三个信源的信源熵较第二个更大。

信源五反映的是二元信源的一般情况，$\delta=0$ 或$\delta=1$ 对应确定事件的分布，因此熵值为 0，而$\delta=0.5$ 对应等概分布，熵值最大。

2．平均条件自信息量（条件熵）

条件熵 $H(X \mid Y)$ 是在二维联合空间 XY 上，对条件自信息量关于 x_i，y_j 取统计平均值得到的计算值，若事件 $x_i y_j$ 的联合分布概率为 $p(x_i y_j)$，给定 y_j 条件下事件 x_i 的条件自信息量为 $I(x_i \mid y_j)$，则 $H(X \mid Y)$ 定义为

$$H(X|Y) \triangleq \sum_i\sum_j p(x_i y_j)I(x_i|y_j) = -\sum_i\sum_j p(x_i y_j)\log\phi(x_i|y_j) \qquad (2\text{-}16)$$

【例 2.10】　一个班级有男生 30 名，女生 10 名，按姓氏笔画排学号，用 y_j 表示学号排序为 1 的学生的性别（y_0 表示男生，y_1 表示女生），用 x_i 表示学号排序为 2 的学生的性别（x_0 表示男生，x_1 表示女生），求 H（Y）和 $H(X \mid Y)$。

学号排序为 1 取男生的概率为 $\omega(y_0)=\frac{30}{40}$，取女生的概率为 $\omega(y_1)=\frac{10}{40}$，则

$$\begin{aligned}H(Y)&=-\omega(y_0)\log\omega(y_0)-\omega(y_1)\log\omega(y_1)\\&=-\frac{30}{40}\log\frac{30}{40}-\frac{10}{40}\log\frac{10}{40}\\&=0.811\text{（比特/符号）}\end{aligned}$$

下面计算 $H(X\mid Y)$，这是一个不放回抽取的例子。若学号 1 是男生，则学号排序为 2 取到男生的概率为 $\phi(x_0|y_0)=\frac{29}{39}$，取到女生的概率为 $\phi(x_1|y_0)=\frac{10}{39}$；若学号 1 是女生，则学号排序为 2 取到男生的概率为 $\phi(x_0|y_1)=\frac{30}{39}$，取到女生的概率为 $\phi(x_1|y_1)=\frac{9}{39}$。

可计算出条件熵，即

$$\begin{aligned}H(X|Y)&=-\sum_{i=0}\sum_{j=0}p(x_iy_j)\log\ \phi(x_i|y_j)\\&=-\sum_{j=0}^{1}\omega(y_j)\sum_{i=0}^{1}\phi(x_i|y_j)\log\phi(x_i|y_j)\\&=-\frac{30}{40}\left(\frac{29}{39}\log\frac{29}{39}+\frac{10}{39}\log\frac{10}{39}\right)-\frac{10}{40}\left(\frac{30}{39}\log\frac{30}{39}+\frac{9}{39}\log\frac{9}{39}\right)\\&=0.811\text{（比特/符号）}\end{aligned}$$

当 X,Y 统计独立时，有 $p(x_iy_j)=q(x_i)\,\omega(y_j)$，$\varphi(x_i\mid y_j)=q(x_i)$，则

$$\begin{aligned}H(X|Y)&=-\sum_j\omega(y_j)\sum_i q(x_i)\log q(x_i)\\&=-\sum_i q(x_i)\log q(x_i)\\&=H(X)\end{aligned}\qquad\text{（2-17）}$$

式（2-17）说明，当 X,Y 相互统计独立时，若集合 $X=\{x_1, x_2, \cdots\}$中所有事件的平均不确定程度为 $H(X)$，则已知 $Y=\{y_1, y_2, \cdots\}$对于确定 $X=\{x_1, x_2, \cdots\}$中的事件没有任何帮助，平均不确定程度 $H(X\mid Y)$仍是 $H(X)$。

从通信角度来看，若将 $X=\{x_1, x_2, \cdots, x_i, \cdots\}$视为信源输出符号，$Y=\{y_1, y_2, \cdots, y_j, \cdots\}$视为信宿接收符号，$I(x_i\mid y_j)$可看做信宿收到 y_j 后，关于发送的符号是否为 x_i 仍然存在的疑义度（不确定性），$H(X|Y)=\sum_i\sum_j p(x_iy_j)I(x_i|y_j)$ 则反映了，经过通信后信宿符号 $y_j(j=1, 2, \cdots)$关于信源符号 $x_i(i=1, 2, \cdots)$ 的平均不确定性。

存在以下两种极端情况：

（1）对于无噪信道 $H(X\mid Y)=0$，信源事件 $X=\{x_1, x_2, \cdots, x_i, \cdots\}$和信宿事件 $Y=\{y_1, y_2, \cdots, y_j, \cdots\}$是一一对应的关系，信宿收到 Y 中的某个元素 y_j 后，关于发送的符号是否为 X 中的某个元素 x_i 不再存在疑义度（不确定性）。

（2）在强噪声情况下，收到的 Y 与 X 毫不相干，可视为统计独立，$H(X\mid Y)=H(X)$。

类似地，若给定 x_i 条件下事件 y_j 的条件自信息量为 $I(y_j\mid x_i)$，则 $H(Y\mid X)$定义为

$$H(Y|X)\ \triangleq\ \sum_i\sum_j p(x_iy_j)I(y_j|x_i)=-\sum_i\sum_j p(x_iy_j)\log p(y_j|x_i)\qquad\text{（2-18）}$$

当 X,Y 统计独立时，有 $p(x_iy_j)=q(x_i)\,\omega(y_j)$，$p(y_j\mid x_i)=\omega(y_j)$，则

$$H\left(Y\middle|X\right)=-\sum_i q(x_i)\sum_j \omega(y_j)\log\omega(y_j)=-\sum_j \omega(y_j)\log\omega(y_j)=H\left(Y\right) \qquad (2\text{-}19)$$

式（2-19）说明，当 X,Y 相互统计独立时，若集合 $Y=\{y_1, y_2, \cdots\}$中所有事件的平均不确定程度为 $H(Y)$，则已知 $X=\{x_1, x_2, \cdots\}$对于确定 $Y=\{y_1, y_2, \cdots\}$中的事件没有任何帮助，平均不确定程度 $H(Y\mid X)$仍是 $H(Y)$。

从通信角度来看，$H(Y\mid X)$是发出确定消息 x_i 后，由于信道干扰而使 y_j 存在的平均不确定性，称 $H(Y\mid X)$为噪声熵（散布度），它表示了由于噪声干扰而使 x_i 错位的一种发散程度。

存在以下两种极端情况：

（1）对于无扰信道，信源事件集 $X=\{x_1, x_2, \cdots, x_i, \cdots\}$和信宿事件集 $Y=\{y_1, y_2, \cdots, y_j, \cdots\}$是一一对应的关系，不会产生错位，$H(Y\mid X)=0$。

（2）对于强噪信道，由于噪声太强，发了 X 等于没发，$H(Y\mid X)$全部由噪声决定，信宿方得不到关于信源 X 的信息，相当于 X 和 Y 相互统计独立，有 $H(Y\mid X)=H(Y)$。

3．联合熵

联合熵 $H(XY)$ 是定义在二维空间 XY 上，对元素 $x_i\ y_j$ 的自信息量的统计平均值，若记事件 $x_i\,y_j$ 出现的概率为 $p(x_i y_j)$，其自信息量为 $I(x_i y_j)$，则联合熵 $H(X\,Y)$ 定义为

$$H\left(XY\right)\ \triangleq\ \sum_i\sum_j p(x_i y_j)I(x_i y_j)=-\sum_i\sum_j p(x_i y_j)\log p(x_i y_j) \qquad (2\text{-}20)$$

式（2-20）可视为 $H(X)$在二维空间 XY 上的推广。

由熵、条件熵、联合熵的定义式（2-15）、式（2-16）、式（2-18）和式（2-20）可导出三者的关系式

$$H(XY)=H(X)+H(Y\mid X)=H(Y)+H(X\mid Y) \qquad (2\text{-}21)$$

式（2-21）反映了信息的可加性。

当 X, Y 统计独立时，有

$$H(XY)=H(X)+H(Y) \qquad (2\text{-}22)$$

【例 2.11】 计算例 1.4 中 Markov 信源的熵。

如前所述，马尔可夫信源每发出一个符号便转入另一状态，马尔可夫信源的数学模型为

$$\begin{bmatrix} s_i\to s_j \\ p(s_j\mid s_i)\end{bmatrix}=\begin{bmatrix} s_0\to s_0 & s_0\to s_0 & s_1\to s_1 & s_1\to s_0 & s_1\to s_1 \\ p(s_0\mid s_0) & p(s_1\mid s_0) & p(s_0\mid s_1) & p(s_1\mid s_1)\end{bmatrix}$$

因此，Markov 信源的熵是条件熵，即

$$H(S)=-\sum_i P(s_i)\sum_j p(s_j\mid s_i)\log p(s_j\mid s_i)$$

例 1.4 中已求出 $\quad p(s_1)=\dfrac{3}{13},\qquad p(s_2)=\dfrac{8}{13},\qquad p(s_3)=\dfrac{2}{13}$

和

$$\begin{cases} p(s_1\mid s_1)=\dfrac{2}{3} & p(s_2\mid s_1)=\dfrac{1}{3} & p(s_3\mid s_1)=0 \\ p(s_1\mid s_2)=0 & p(s_2\mid s_2)=\dfrac{3}{4} & p(s_3\mid s_2)=\dfrac{1}{4} \\ p(s_1\mid s_3)=\dfrac{1}{2} & p(s_2\mid s_3)=\dfrac{1}{2} & p(s_3\mid s_3)=0 \end{cases}$$

由此可算出此 Markov 信源的熵为

$$
\begin{aligned}
H = &-p(s_1)[p(s_1 \mid s_1)\log p(s_1 \mid s_1) + p(s_2 \mid s_1)\log p(s_2 \mid s_1)] - \\
&p(s_2)[p(s_2 \mid s_2)\log p(s_2 \mid s_2) + p(s_3 \mid s_2)\log p(s_3 \mid s_2)] - \\
&p(s_3)[p(s_1 \mid s_3)\log p(s_1 \mid s_3) + p(s_2 \mid s_3)\log p(s_2 \mid s_3)] \\
= &-\frac{3}{13}\left(\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right)-\frac{8}{13}\left(\frac{3}{4}\log\frac{3}{4}+\frac{1}{4}\log\frac{1}{4}\right)-\frac{2}{13}\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right) \\
= &\ 0.865\text{（比特/符号）}
\end{aligned}
$$

2.2.2 熵函数的性质

1. 凸集合与凸函数

本章的定理证明多次用到凸函数的概念，几个重要的物理量也与凸函数概念有关。例如，信源熵 $H(X)$是信源分布 $q(x)$的∩型凸函数；平均互信息量 $I(X;Y)$是信源分布 $q(x)$的∩型凸函数，是信道转移概率 $p(y|x)$的∪型凸函数；率失真函数 $R(D)$是 D 的∪型凸函数。

下面简单介绍凸集和凸函数的概念。

定义 2.1 令 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 和 $\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_n)$ 是 n 维实矢量空间集合 $\boldsymbol{R}$ 中任意两个 n 维矢量，对实数 θ，$0\leqslant\theta\leqslant1$，有

$$\theta\boldsymbol{\alpha}+(1-\theta)\boldsymbol{\beta}\in\boldsymbol{R}$$

则称 $\boldsymbol{R}$ 为凸集合。

从几何上来看，若 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 是集合 $\boldsymbol{R}$ 中的任意两点，$\theta\boldsymbol{\alpha}+(1-\theta)\boldsymbol{\beta}$ 表示这两点间的连线，若该连线也在集合 $\boldsymbol{R}$ 中，则称为 $\boldsymbol{R}$ 凸集。下面给出几个凸集和非凸集合的例子。

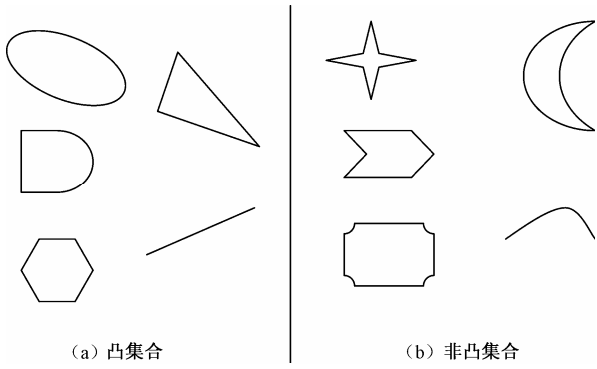


图 2-3　一维、二维凸集合和非凸集合的例子

（1）∩型凸函数

定义 2.2 设 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 为一个 n 元函数，若对任意 $f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\in f(\boldsymbol{x})$，任意正数 θ，$0\leqslant\theta\leqslant1$，有

$$\theta f(\boldsymbol{x}_1)+(1-\theta)f(\boldsymbol{x}_2)\leqslant f[\theta\boldsymbol{x}_1+(1-\theta)\boldsymbol{x}_2] \qquad (2\text{-}23)$$

则称 $f(\boldsymbol{x})$为定义域上的∩型凸函数，式（2-23）可理解为“函数的均值≤均值的函数”。

一元∩型凸函数可用图 2-4 所示的几何图形表示。

（2）∪型凸函数

定义 2.2 设 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 为一个 n 元函数，若对任意 $f(\boldsymbol{x}_1),f(\boldsymbol{x}_2)\in f(\boldsymbol{x})$，任意正数 θ，$0\leqslant\theta\leqslant1$，有

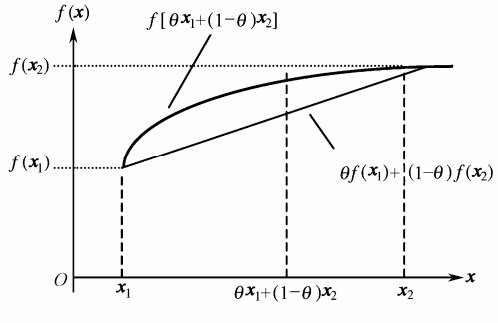


图 2-4　一元∩型凸函数

$$f[\theta \boldsymbol{x}_1+(1-\theta)\boldsymbol{x}_2] \leqslant \theta f(\boldsymbol{x}_1)+(1-\theta) f(\boldsymbol{x}_2) \tag{2-24}$$

则称 $f(\boldsymbol{x})$为定义域上的∪型凸函数，式（2-24）可理解为“均值的函数≤函数的均值”。

一元∪型凸函数可用图 2-5 所示的几何图形表示。

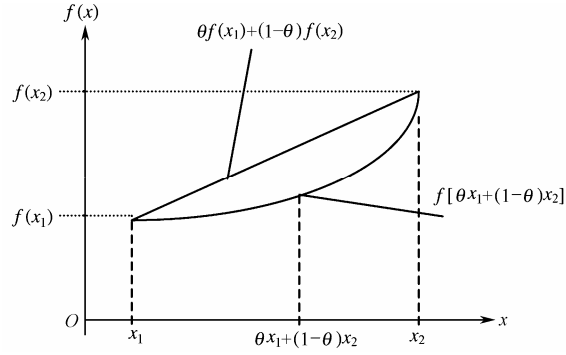


图 2-5　一元∪型凸函数

2．极大离散熵定理

设信源的消息个数为 M，则 $H(X) \leqslant \log M$，等号当且仅当信源 X 中各消息等概$\left(=\dfrac{1}{M}\right)$时成立，即各消息等概分布时，信源熵最大。

证明方法一：利用不等式 $\ln x \leqslant x-1$，等号在 $x=1$ 时成立，如图 2-6 所示。

注意，$\log x = \ln x \log \mathrm{e}$。

$$\begin{aligned} H(X)-\log M &= \sum_i q(x_i)\log\frac{1}{q(x_i)} - \sum_i q(x_i)\log M \\ &= \sum_i q(x_i)\log\frac{1}{q(x_i)M} \\ &= \sum_i q(x_i)\ln\frac{1}{q(x_i)M}\log \mathrm{e} \\ &\leqslant \sum_i q(x_i)\left(\frac{1}{q(x_i)M}-1\right)\log \mathrm{e} \end{aligned}$$

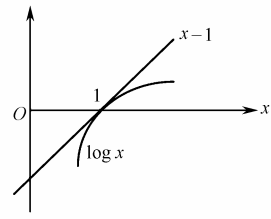


图 2-6　ln$x$≤x–1 关系曲线

$$=\left(\sum_i \frac{1}{M}-\sum_i q(x_i)\right)\log \mathrm{e}$$
$$=(1-1)\log \mathrm{e}=0$$

证毕

证明方法二：利用 $\log x$ 的∩型凸函数性质，可得

$$H(X)-\log M=\sum_i q(x_i)\log\frac{1}{q(x_i)M}$$
$$\leq \log\sum_i q(x_i)\frac{1}{q(x_i)M}$$
$$=\log\sum_i \frac{1}{M}$$
$$=\log 1=0$$

证毕

上面两种证明方法是信息论中经常用到的证明方法。

如前所述，1 比特信息是一位二进制数取“0”或“1”等概分布时所提供的信息量，而最大离散熵定理说明等概分布时信息熵最大，所以，实际上 1 比特信息就是一位二进制数所能提供的最大信息量。

3．熵函数的性质

（1）对称性

集合 $X=\{x_1, x_2, \cdots, x_N\}$中的各元素 $x_1, x_2, \cdots, x_N$ 任意改变顺序时，熵只和分布（概率）有关，而不关心某个具体事件对应哪个概率。

例如，$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$ 和 $\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/8 & 1/8 & 1/4 & 1/2 \end{bmatrix}$ 的熵是相等的。

（2）非负性

$H(X)\geqslant 0$。因为 $H(X)=\sum_{x_i} q(x_i)I(x_i)$，由式（2-2）知 $I(x_i)=-\log q(x_i)\geqslant 0$，因此有

$$H(X)\geqslant 0$$

（3）确定性

在集合 $X=(x_1, x_2, \cdots, x_N)$中，若有一个事件 x_i（i=1，2，…，N）是必然事件，则其余事件必为不可能事件，即该集合的概率分布为 $\begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots & x_N \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}$，根据式（2-15）有 $H(X)=-\sum_i q(x_i)\log q(x_i)$，可以计算得 $H(X)=0$。

（4）扩展性

离散事件集 $\begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ p_1 & p_2 & \cdots & p_N \end{bmatrix}$，增加一个不可能事件 x_{N+1} 后，得到集合 $\begin{bmatrix} x_1 & x_2 & \cdots & x_N & x_{N+1} \\ p_1 & p_2 & \cdots & p_N-\delta & \delta \end{bmatrix}$，$\delta\to 0$，则两个集合的熵相等，即事件 x_{N+1} 的概率与事件集 X 中其他事件的概率相比很小，它对集合熵值的贡献可以忽略不计。

（5）可加性

集合 $X=\{x_1, x_2, \cdots, x_i, x_{i+1}, \cdots, x_N\}$的概率分布为

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_i & x_{i+1} & \cdots & x_N \\ p(x_1) & p(x_2) & \cdots & p(x_i) & p(x_{i+1}) & \cdots & p(x_N) \end{bmatrix}$$

则下式成立：

$$H(X) = H(x_1，\ x_2，\ \cdots，\ x_i，\ x_{i+1}，\ \cdots，\ x_N)$$

$$= H(x_1, x_2, \cdots, x_{i-1}, x_i + x_{i+1}, x_{i+2}, \cdots, x_N) + (p_i + p_{i+1}) H(\frac{p_i}{p_i + p_{i+1}}, \frac{p_{i+1}}{p_i + p_{i+1}}) \quad (2\text{-}25)$$

式（2-25）说明，若原集合 X 的熵为 $H(X)$，则对将集合 X 中的两个事件 x_i 和 x_{i+1} 合并成一个事件 $x_i + x_{i+1}$ 后得到的含有 $N-1$ 个事件的集合，有

$$X^{(N-1)} = \{x_1，\ x_2，\ \cdots，\ x_i + x_{i+1}，\ x_{i+2}，\ \cdots，\ x_N\}$$

计算集合 $X^{(N-1)}$ 的熵，结果为 $N-1$ 个事件集合 $X^{(N-1)}$ 的熵再加上式（2-25）中的第二项。第二项表明，事件 $x_i + x_{i+1}$ 发生后，到底是事件 x_i 发生还是事件 x_{i+1} 发生仍存在不确定性，在事件 $x_i + x_{i+1}$ 发生的条件下，这两个事件必有一个发生，x_i 发生的概率为 $\frac{p_i}{p_i + p_{i+1}}$，$x_{i+1}$ 发生的概率为 $\frac{p_{i+1}}{p_i + p_{i+1}}$。

【例 2.12】 集合 $X = \{x_1，\ x_2，\ x_3，\ x_4，\ x_5\}$ 的概率分布为 $\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 1/4 & 1/4 & 1/8 & 1/8 & 1/4 \end{bmatrix}$，若将 x_1 和 x_2 合并成一个事件，得到新的集合 $X^{(4)} = \{x_1 + x_2，\ x_3，\ x_4，\ x_5\}$，则其概率分布为 $\begin{bmatrix} x_1 + x_2 & x_3 & x_4 & x_5 \\ 1/2 & 1/8 & 1/8 & 1/4 \end{bmatrix}$。

分别计算两个集合的熵，得

$$H(X) = -3 \times \frac{1}{4} \log \frac{1}{4} - 2 \times \frac{1}{8} \log \frac{1}{8} = 2.25 \text{（比特/符号）}$$

$$H(X^{(4)}) = -\frac{1}{2} \log \frac{1}{2} - 2 \times \frac{1}{8} \log \frac{1}{8} - \frac{1}{4} \log \frac{1}{4} = 1.75 \text{（比特/符号）}$$

另外可以算得

$$p(x_1 + x_2) H(x_1, x_2) = \frac{1}{2} \left(-\frac{1/4}{1/4 + 1/4} \log \frac{1/4}{1/4 + 1/4} - \frac{1/4}{1/4 + 1/4} \log \frac{1/4}{1/4 + 1/4} \right) = 0.5 \text{（比特/符号）}$$

可见，$H(X) = H(X^{(4)}) + p(x_1 + x_2) H(x_1, x_2)$ 成立。

（6）条件熵小于等于无条件熵，即

$$H(X \mid Y) \leqslant H(X)$$

X,Y 统计独立时等号成立。

通俗地说，条件化使不确定性降低，从而熵减小。

证明： 因为

$$H(X \mid Y) - H(X) = \sum_i \sum_j p(x_i y_j) \log \frac{1}{\phi(x_i \mid y_j)} - \sum_i q(x_i) \log \frac{1}{q(x_i)}$$

$$= \sum_i \sum_j p(x_i y_j) \log \frac{q(x_i)}{\phi(x_i \mid y_j)}$$

$$\leq \sum_i \sum_j p(x_i y_j)\left(\frac{q(x_i)}{\phi(x_i|y_j)}-1\right)\log \mathrm{e}$$

$$=\left(\sum_i \sum_j q(x_i)\omega(y_j)-1\right)\log \mathrm{e}=0$$

所以 $H(X|Y)\leqslant H(X)$，当 $\dfrac{q(x)}{\phi(x|y)}=1$（即 X,Y 统计独立）时等号成立。　　证毕

（7）联合熵大于等于独立事件的熵，小于等于两独立事件熵之和，即

$$\begin{cases} H(XY)\geq H(X) \\ H(XY)\geq H(Y) \end{cases} \tag{2-26}$$

$$H(XY)\leqslant H(X)+H(Y) \tag{2-27}$$

证明式（2-26）：根据熵的非负性 $H(Y|X)\geqslant 0$，由式（2-21）得

$$H(XY)=H(X)+H(Y|X)\geqslant H(X)$$

同样有　　$$H(XY)=H(Y)+H(X|Y)\geqslant H(Y)$$

证明式（2-27）：根据熵的第 6 条性质，条件熵小于或等于无条件熵，有

$$H(XY)=H(X)+H(Y|X)\leqslant H(X)+H(Y)$$　　证毕

【例 2.13】　某一平稳信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 & 2 \\ 11/36 & 16/36 & 9/36 \end{bmatrix}$，若此信源发出的符号只与前一符号有关，用联合概率给出它们之间的关联程度，如表 2-5 所示，并满足 $\sum_i \sum_j p(x_i x_j)=1$。

表 2-5　信源的联合概率密度 $p(x_i x_j)$

x_j \ x_i	0	1	2
0	1/4	1/18	0
1	1/18	1/3	1/18
2	0	1/18	7/36

根据概率关系式 $p(x_i x_j)=p(x_j|x_i)\,q(x_i)$，可算出条件概率 $p(x_j|x_i)$，如表 2-6 所示。

表 2-6　信源的条件概率密度 $p(x_j|x_i)$

x_j \ x_i	0	1	2
0	9/11	1/8	0
1	2/11	3/4	2/9
2	0	1/8	7/9

并满足 $\sum_j p(x_j|x_i)=1$。

若认为各符号之间无关联，则可算出信源熵

$$H(X)=\sum_i q(x_i)\log q(x_i)$$

$$=-\frac{11}{36}\log\frac{11}{36}-\frac{16}{36}\log\frac{16}{36}-\frac{9}{36}\log\frac{9}{36}$$

$$=1.54\text{（比特/符号）}$$

考虑各符号之间有关联，可算出条件熵

$$H(X_2|X_1)=-\sum_i\sum_j p(x_ix_j)\log p(x_j|x_i)$$

$$=-\frac{1}{4}\log\frac{9}{11}-\frac{1}{18}\log\frac{2}{11}-\frac{1}{18}\log\frac{1}{8}-\frac{1}{3}\log\frac{3}{4}-\frac{1}{18}\log\frac{1}{8}-\frac{1}{18}\log\frac{2}{9}-\frac{7}{36}\log\frac{7}{9}$$

$$=0.87\text{（比特/符号）}$$

联合熵

$$H(X_1X_2)=-\sum_i\sum_j p(x_ix_j)\log p(x_ix_j)$$

$$=-\frac{1}{4}\log\frac{1}{4}-\frac{4}{18}\log\frac{1}{18}-\frac{1}{3}\log\frac{1}{3}-\frac{7}{36}\log\frac{7}{36}=2.41\text{（比特/符号）}$$

由此可见，$H(X_1X_2)=H(X_1)+H(X_2|X_1)$。

因为 $H(X_2|X_1)<H(X_1)$，所以有 $H(X_1X_2)<2H(X_1)$。

信源的条件熵小于无条件熵，正是由于符号间的关联使得信源的不确定程度减少，从而使熵减少。

2.3 离散集的平均互信息量

2.3.1 平均互信息量

前面研究的是单个事件的信息量及单事件集的平均自信息量，这是最简单的离散信源，如果将发送符号与接收符号看成两个不同的“信源”，则通过信道的转移概率将二者统计结合起来，就可以来讨论信息的流通问题，那么一次通信从发送到接收究竟能得到多少信息量呢，这就是本节要讨论的平均互信息量。

1．平均互信息量

前面已定义 $x_i \in X$ 和 $y_j \in Y$ 之间的互信息量为 $I(x_i;y_j)$，在集合 X 上对 $I(x_i;y_j)$进行概率加权统计平均，可得 $I(X;y_j)$为

$$I(X;y_j)=\sum_i\phi(x_i|y_j)\cdot I(x_i;y_j)$$

$$=\sum_i\phi(x_i|y_j)\log\frac{\phi(x_i|y_j)}{q(x_i)} \qquad (2\text{-}28)$$

式（2-28）是特定事件 $y_j \in Y$ 出现时，所给出的关于事件集 $X=(x_1,\cdots,x_i,\cdots)$的平均互信息量。根据式（2-28）有

$$-I(X;y_j)=\sum_i\phi(x_i|y_j)\log\frac{q(x_i)}{\phi(x_i|y_j)}$$

$$\le\log\sum_i\phi(x_i|y_j)\frac{q(x_i)}{\phi(x_i|y_j)}$$

$$=\log\sum_i q(x_i)=0$$

即

$$I(X;y_j)\geqslant 0 \tag{2-29}$$

由互信息量的性质（4）知道，$I(x_i;y_j)=\log\dfrac{\phi(x_i|y_j)}{q(x_i)}$ 的取值可能为正值，也可能为负值，这是因为将 y_j 视为一个特定的接收符号，x_i 视为信源的任一消息，由于后验概率 $\phi(x_i|y_j)$ 有可能大于 $q(x_i)$，也有可能小于先验概率 $q(x_i)$，所以 $I(x_i;y_j)$ 是一个可正可负的物理量。而式（2-29）说明接收到 y_j 后，提供的关于信源所有消息的平均互信息量总是非负的，总是有利于做出判决。

再将式（2-28）对集合 Y 进行统计平均，就可以得到平均互信息量

$$\begin{aligned} I(X;Y)&=\sum_j \omega(y_j)I(X;y_j)\\ &=\sum_i\sum_j \omega(y_j)\phi(x_i|y_j)\log\frac{\phi(x_i|y_j)}{q(x_i)}\\ &=\sum_i\sum_j p(x_iy_j)\log\frac{\phi(x_i|y_j)}{q(x_i)}\\ &=\sum_i\sum_j p(x_iy_j)I(x_i;y_j)\end{aligned}$$

即有

$$\begin{aligned} I(X;Y)&=\sum_i\sum_j p(x_iy_j)I(x_i;y_j)\\ &=\sum_i\sum_j p(x_iy_j)\log\frac{p(x_iy_j)}{q(x_i)\omega(y_j)}\end{aligned} \tag{2-30}$$

当 X,Y 统计独立时，$I(x_i;y_j)=0$，从而 $I(X;Y)=0$。

【例 2.14】 二元等概信源 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_0 & x_1\\ 1/2 & 1/2\end{bmatrix}$，通过信道转移概率为 $\boldsymbol{P}=\begin{matrix}x_0\\ x_1\end{matrix}\begin{matrix}\overset{y_0}{\overset{\downarrow}{}}\\ \overset{y_1}{\overset{\downarrow}{}}\end{matrix}\begin{bmatrix}5/6 & 1/6\\ 1/2 & 1/2\end{bmatrix}$ 的信道传输，信宿接收符号 $Y=\{y_0,y_1\}$，计算信源与信宿间的平均互信息量 $I(X;Y)$。

（1）先根据 $\omega(y_j)=\sum_i p(y_j|x_i)q(x_i)$ 计算出

$$\omega(y_0)=p(y_0|x_0)q(x_0)+p(y_0|x_1)q(x_1)=\frac{1}{2}\left(\frac{5}{6}+\frac{1}{2}\right)=\frac{2}{3}$$

$$\omega(y_1)=p(y_1|x_0)q(x_0)+p(y_1|x_1)q(x_1)=\frac{1}{2}\left(\frac{1}{6}+\frac{1}{2}\right)=\frac{1}{3}$$

（2）计算各消息之间的互信息量 $I(x_i;y_j)$

$$I(x_0;y_0)=\log\frac{p(y_0|x_0)}{w(y_0)}=\log\frac{5/6}{2/3}=\log\frac{5}{4}=0.322\ \text{（比特）}$$

$$I(x_0;y_1)=\log\frac{p(y_1|x_0)}{w(y_1)}=\log\frac{1/6}{1/3}=\log\frac{1}{2}=-1\ \text{（比特）}$$

$$I(x_1;y_0)=\log\frac{p(y_0|x_1)}{w(y_0)}=\log\frac{1/2}{2/3}=\log\frac{3}{4}=-0.415\ \text{（比特）}$$

$$I(x_1;y_1)=\log\frac{p(y_1\mid x_1)}{w(y_1)}=\log\frac{1/2}{1/3}=\log\frac{3}{2}=0.585\ \text{（比特）}$$

从计算结果看，$I(x_0;y_1)$和$I(x_1;y_0)$为负值，这是由于信道干扰使后验概率$\phi(x_0\mid y_1)<q(x_0)$和$\phi(x_1\mid y_0)<q(x_1)$，也就是在这两种情况下，通信后关于发送的是哪个符号，其疑义度$I(x_i\mid y_j)=-\log p(x_i\mid y_j)$大于发送前的疑义度$I(x_i)=-\log q(x_i)$，这不利于做出判决，使一次通信后得到的信息量反而为负值。

（3）计算平均互信息量

$$\begin{aligned}I(X;Y)&=\sum_i\sum_j p(x_iy_j)\,I(x_i;y_j)\\&=\sum_i\sum_j q(x_i)p(y_j\mid x_i)I(x_i;y_j)\\&=\frac{1}{2}\left[\frac{5}{6}\times0.322+\frac{1}{6}\times(-1)+\frac{1}{2}\times(-0.415)+\frac{1}{2}\times0.585\right]\\&=0.093\ \text{（比特）}\end{aligned}$$

2．平均条件互信息量

平均条件互信息量$I(X;Y\mid Z)$是在联合概率空间$\{XYZ,\ p(xyz)\}$上定义的物理量。由式（2-11）可知，$I(x_i;y_j\mid z_k)=\log\dfrac{p(x_i\mid y_jz_k)}{p(x_i\mid z_k)}=\log\dfrac{p(x_iy_j\mid z_k)}{p(x_i\mid z_k)p(y_j\mid z_k)}$，对该式在三维空间$XYZ$上求概率加权平均值，就得到平均条件互信息量

$$\begin{aligned}I(X;Y\mid Z)&\triangleq\sum_i\sum_j\sum_k p(x_iy_jz_k)I(x_iy_j\mid z_k)\\&=\sum_i\sum_j\sum_k p(x_iy_jz_k)\log\frac{p(x_iy_j\mid z_k)}{p(x_i\mid z_k)p(y_j\mid z_k)}\end{aligned}\qquad(2\text{-}31)$$

式中，$p(x_iy_jz_k)$满足$\sum_i\sum_j\sum_k p(x_iy_jz_k)=1$。

2.3.2　平均互信息量的性质

1．平均互信息量的性质

（1）非负性

$$\begin{aligned}I(X;Y)&\geq 0\\I(X;Y\mid Z)&\geq 0\end{aligned}\qquad(2\text{-}32)$$

证明：对式（2-29）$I(X;y_j)\geqslant0$两边关于集Y进行统计平均，得

$$\sum_j\omega(y_j)I(X;y_j)=I(X;Y)\geq 0$$

等号在集X和Y统计独立时成立。

（2）互易性

$$I(X;Y)=I(Y;X)\qquad(2\text{-}33)$$

这由定义$I(X;Y)=\sum_i\sum_j p(x_iy_j)\log\dfrac{p(x_iy_j)}{q(x_i)\omega(y_j)}$的对称性可以看出。

（3）$\begin{cases} I(X;Y) \le H(X) & (2\text{-}34a) \\ I(X;Y) \le H(Y) & (2\text{-}34b) \end{cases}$

式（2-34a）说明，在一般情况下，由于存在噪声，经一次通信所得到的信息量 $I(X;Y)$总是小于信源自身所携带的信息量，式（2-34a）和式（2-34b）可分别根据式（2-35）、式（2-36）$\begin{cases} I(X;Y)=H(X)-H(X|Y) \\ I(X;Y)=H(Y)-H(Y|X) \end{cases}$，由熵的非负性、以及熵函数的性质（6）（条件熵小于等于无条件熵）得到。

当 X,Y 完全相关时，X 和 Y 是一一对应的关系，$H(X|Y)=H(Y|X)=0$，式（2-34）中等号成立，说明集 X 可提供给集 Y 的最大信息量为 $H(X)$，而集 Y 可提供给集 X 的最大信息量为 $H(Y)$。

2．平均互信息量与信源熵、条件熵的关系

$$I(X;Y) = H(X)-H(X \mid Y) \tag{2-35}$$

$$I(X;Y) = H(Y)-H(Y \mid X) \tag{2-36}$$

$$I(X;Y) = H(X)+H(Y)-H(XY) \tag{2-37}$$

上面三个等式可以通过分别对式（2-6）、式（2-7）、式（2-8）求统计平均值 $I(X;Y)=\sum_i\sum_j p(x_iy_j)I(x_i;y_j)$ 得到。

$$I(x_i;y_j)=\log\frac{\phi(x_i|y_j)}{q(x_i)}=\log\frac{p(y_j|x_i)}{\omega(y_j)}=\log\frac{p(x_iy_j)}{q(x_i)\omega(y_j)}$$

它们之间的关系可用图 2-7 表示，该图称为维拉图。

下面从通信的角度来讨论平均互信息量 $I(X;Y)$的物理意义。

（1）由第一等式 $I(X;Y)=H(X)-H(X \mid Y)$看 $I(X;Y)$的物理意义

设 X 为发送消息符号集，Y 为接收符号集，$H(X)$是输入集的平均不确定性，$H(X \mid Y)$是观察到 Y 后，集 X 还保留的不确定性（由于存在噪声的原因），二者之差 $I(X;Y)$就是在接收过程中得到的关于 X,Y 的平均互信息量。$H(X \mid Y)$常称为含糊度（疑义度），显然，含糊度越大，通信过程中得到的信息量就越少。

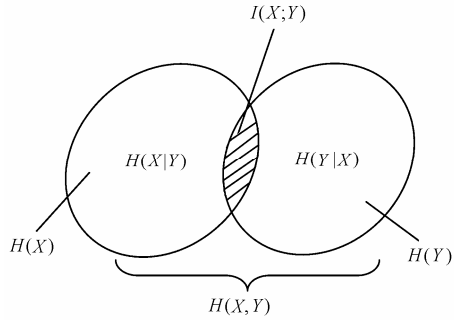


图 2-7　维拉图

对于无扰信道，X 和 Y 是一一对应的关系，观察到 Y 后，对集 X 不再存在疑义，因此有 $H(X \mid Y)=0$，从而 $I(X;Y)=H(X)$。

对于强噪信道，信号全部被噪声所淹没，以至于收到 Y 后不能提供关于 X 的任何信息，

收到 Y 后对 X 保留的不确定性 $H(X|Y)$仍是 X 原来自带的不确定性 $H(X)$，即 $H(X|Y)=H(X)$，从而 $I(X;Y)=0$。

（2）由第二等式 $I(X;Y)=H(Y)-H(Y|X)$看 $I(X;Y)$的物理意义

$H(Y)$是观察到 Y 所获得的信息量，$H(Y|X)$是发出确定消息 X 后，由于干扰而使 Y 存在的平均不确定性，二者之差 $I(X;Y)$就是一次通信所获得的信息量。称 $H(Y|X)$为噪声熵（散布度），它表示由于噪声干扰而使 X 错位的一种发散程度，显然，噪声熵越大，得到的信息量就越小。

对于无扰信道，X 和 Y 是一一对应的关系，不会产生错位，因此 $H(Y|X)=0$，一次通信所获得的信息量就等于观察到 Y 所获得的信息量 $H(Y)$，也就等于通信前 X 自带的信息量 $H(X)$，即有 $I(X;Y)=H(X)=H(Y)$。

对于强噪信道，由于噪声太强，发了 X 等于没发，$H(Y|X)$全部由噪声所决定，信宿方得不到关于 X 的信息，相当于 X 和 Y 相互统计独立，有 $H(Y|X)=H(Y)$，从而 $I(X;Y)=0$。

（3）由第三等式 $I(X;Y)=H(X)+H(Y)-H(X,Y)$ 看 $I(X;Y)$的物理意义

通信前，随机变量 X 和随机变量 Y 可视为统计独立，其先验不确定性为 $H(X)+H(Y)$，信道的统计特性由信道转移概率 $p(y|x)$描述，通信后，在信道两端出现的随机变量 X 和 Y 由于 $p(y|x)$有了统计联系，整个系统的后验不确定性为 $H(XY)$，二者之差 $H(X)+H(Y)-H(XY)$就是通信过程中不确定性减少的量，也就是通信过程中获得的平均互信息量 $I(X;Y)$。

【例 2.15】 已知信源消息集为 $X=\{0,1\}$，接收符号集为 $Y=\{0,1\}$，通过有扰信道传输，其传输特性如图 2-8 所示，这是一个二进制对称信道 BSC。已知先验概率 $q(0)=q(1)=\dfrac{1}{2}$，计算平均互信息量 $I(X;Y)$及各种熵。

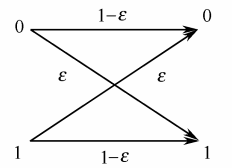


图 2-8　二进制对称信道

记 $q(x)$为信源输入概率，$\omega(y)$为信宿输出概率，$p(y|x)$为信道转移概率，$\phi(x|y)$为后验概率。

（1）由图 2-8 得 $\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}$，先算出 $p(x_iy_j)=q(x_i)\,p(y_j|x_i)$，即

$$p(00)=q(0)p(0|0)=0.5(1-\varepsilon)\qquad p(01)=q(0)p(1|0)=0.5\varepsilon$$
$$p(10)=q(1)p(0|1)=0.5\varepsilon\qquad p(11)=q(1)p(1|1)=0.5(1-\varepsilon)$$

（2）计算 $\omega(y_j)=\sum_i p(x_iy_j)$，得

$$\omega(0)=\sum_i p(x_i0)=p(00)+p(10)=0.5(1-\varepsilon)+0.5\varepsilon=0.5$$
$$\omega(1)=\sum_i p(x_i1)=p(01)+p(11)=0.5\varepsilon+0.5(1-\varepsilon)=0.5$$

计算后验概率，得

$$\phi(0|0)=\frac{p(00)}{\omega(0)}=\frac{0.5(1-\varepsilon)}{0.5}=1-\varepsilon\qquad \phi(0|1)=\frac{p(01)}{\omega(1)}=\frac{0.5\varepsilon}{0.5}=\varepsilon$$
$$\phi(1|0)=\frac{p(10)}{\omega(0)}=\frac{0.5\varepsilon}{0.5}=\varepsilon\qquad \phi(1|1)=\frac{p(11)}{\omega(1)}=\frac{0.5(1-\varepsilon)}{0.5}=1-\varepsilon$$

（4）计算各种熵及平均互信息量

信源熵 $H(X)=-\sum_i q(x_i)\log q(x_i)=-0.5\log 0.5-0.5\log 0.5=\log 2$

信宿熵 $H(Y)=-\sum_j \omega(y_j)\log\omega(y_j)=-0.5\log 0.5-0.5\log 0.5=\log 2$

联合熵 $H(XY)=-\sum_i\sum_j p(x_iy_j)\log p(x_iy_j)$

$$=-2\times 0.5\,(1-\varepsilon)\log 0.5(1-\varepsilon)-2\times 0.5\varepsilon\log 0.5\varepsilon$$
$$=\log 2-(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon$$
$$=\log 2+H_2(\varepsilon)$$

式中，$H_2(\varepsilon)\triangleq -(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon$。

平均互信息量 $I(X;Y)=H(X)+H(Y)-H(XY)=\log 2+H_2(\varepsilon)$

可疑度 $H(X|Y)=-\sum_i\sum_j p(x_iy_j)\log\phi(x_i|y_j)$

$$=-p(00)\log\phi(0\mid 0)-p(01)\log\phi(0\mid 1)-p(10)\log\phi(1\mid 0)-p(11)\log\phi(1\mid 1)$$
$$=-2\times 0.5\,(1-\varepsilon)\log(1-\varepsilon)-2\times 0.5\varepsilon\log\varepsilon=H_2(\varepsilon)$$

散布度 $H(Y|X)=-\sum_i\sum_j p(x_iy_j)\log p(y_j|x_i)$

$$=-p(00)\log p(0\mid 0)-p(01)\log p(1\mid 0)-p(10)\log p(0\mid 1)-p(11)\log p(1\mid 1)$$
$$=-2\times 0.5\,(1-\varepsilon)\log(1-\varepsilon)-2\times 0.5\varepsilon\log\varepsilon=H_2(\varepsilon)$$

此例的计算结果是对维拉图标示的各种熵及平均互信息量之间关系的验证。

2.3.3 有关平均互信息量的两条定理

研究通信问题，主要研究的是信源和信道，它们的统计特性可以分别用消息先验概率 $q(x)$ 及信道转移概率 $p(y\mid x)$来描述，而平均互信息量 $I(X;Y)$是经过一次通信后信宿所获得的信息。

由式（2-30）知，平均互信息量定义为

$$I(X;Y)=\sum_i\sum_j p(x_iy_j)\log\frac{p(x_iy_j)}{q(x_i)\omega(y_j)}$$
$$=\sum_i\sum_j p(y_j|x_i)q(x_i)\log\frac{p(y_j|x_i)q(x_i)}{q(x_i)\sum_i p(x_iy_j)}$$
$$=\sum_i\sum_j p(y_j|x_i)q(x_i)\log\frac{p(y_j|x_i)}{\sum_i p(y_j|x_i)q(x_i)} \qquad (2\text{-}38)$$

式（2-38）说明，$I(X;Y)$是信源分布概率 $q(x)$和信道转移概率 $p(y\mid x)$的函数，下面两条定理阐明了 $I(X;Y)$与 $q(x)$和 $p(y\mid x)$之间的关系。

定理 2.1 当信道给定，即信道转移概率 $p(y\mid x)$固定时，平均互信息量 $I(X;Y)$是信源概率分布 $q(x)$的∩型凸函数。

根据式（2-23），在信道固定的情况下，如果给定两个信源分布 $q_1(x)$和 $q_2(x)$，分别对应平均互信息量 $I_1(X;Y)$和 $I_2(X;Y)$，记概率分布 $q(x)=\theta q_1(x)+(1-\theta)q_2(x)$（式中 $0\leqslant\theta\leqslant 1$），对应平均互信息量 $I(X;Y)$，若 $I(X;Y)$是∩型凸函数，则应满足

$$\theta I_1(X;Y)+(1-\theta)I_2(X;Y)\leqslant I(X;Y) \qquad (2\text{-}39)$$

式（2-39）表示，函数的均值小于等于均值的函数，如图 2-9 所示。

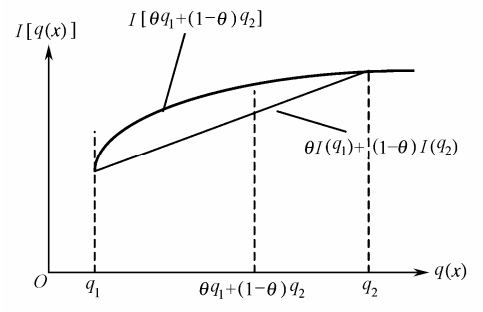


图 2-9　函数的均值小于等于均值的函数

证明：

$$\begin{aligned}
&\theta I_1(X;Y)+(1-\theta)I_2(X;Y)-I(X;Y)\\
&=\theta\sum_i\sum_j p_1(x_iy_j)\log\frac{p(y_j|x_i)}{\omega_1(y_j)}+(1-\theta)\sum_i\sum_j p_2(x_iy_j)\log\frac{p(y_j|x_i)}{\omega_2(y_j)}-\\
&\quad\sum_i\sum_j[\theta\cdot p_1(x_iy_j)+(1-\theta)p_2(x_iy_j)]\log\frac{p(y_j|x_i)}{\omega(y_j)}\\
&=\theta\sum_i\sum_j p_1(x_iy_j)\log\frac{\omega(y_j)}{\omega_1(y_j)}+(1-\theta)\sum_i\sum_j p_2(x_iy_j)\log\frac{\omega(y_j)}{\omega_2(y_j)}\\
&\le\ \theta\log\sum_i\sum_j p_1(x_iy_j)\frac{\omega(y_j)}{\omega_1(y_j)}+(1-\theta)\log\sum_i\sum_j p_2(x_iy_j)\frac{\omega(y_j)}{\omega_2(y_j)}\\
&=\theta\log\sum_j\frac{\omega_1(y_j)\omega(y_j)}{\omega_1(y_j)}+(1-\theta)\log\sum_j\frac{\omega_2(y_j)\omega(y_j)}{\omega_2(y_j)}\\
&=\theta\log1+(1-\theta)\log1=0
\end{aligned}$$

所以

$$\theta I_1(X;\ Y)\ +\ (1-\theta)\ I_2\ (X;\ Y)\leqslant \mathrm{I}\ (X;\ Y)$$

证毕

定理 2.1 说明，当信道固定时，对于不同的信源分布，信道输出端获得的信息量是不同的。因此，可以预见，对于每一个固定信道，一定存在一种信源（一种分布）$q(x)$，使输出端获得的信息量最大，因为∩型凸函数存在最大值。

【例 2.16】　二进制对称信道 BSC 如图 2-10 所示，输入符号集 $X=\{x_1, x_2\}=\{0,1\}$，输出符号集 $Y=\{y_1,y_2\}=\{0,1\}$，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}$，信源概率分布 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}0 & 1\\ \delta & 1-\delta\end{bmatrix}$，计算平均互信息量 $I(X;Y)=H(Y)-H(Y\mid X)$。

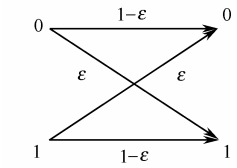


图 2-10　二进制对称信道

解：先由 $\omega(y_j)=\sum_i q(x_i)\,p(y_j|x_i)$ 算出

$$\omega(0)=q(0)\,p(0\mid 0)+q(1)\,p(0\mid 1)=\delta(1-\varepsilon)+(1-\delta)\varepsilon$$

$$\omega(1)=q(0)\,p(1\mid 0)+q(1)\,p(1\mid 1)=\delta\varepsilon+(1-\delta)(1-\varepsilon)=1-[\delta(1-\varepsilon)+(1-\delta)\varepsilon]=1-\omega(0)$$

再计算熵和条件熵

$$H(Y)=-\sum_{j=1}^{2}\omega(y_j)\log\omega(y_j)$$

$$=-\omega(0)\log\omega(0)-[1-\omega(0)]\log[1-\omega(0)]$$

$$=H_2[\omega(0)]$$

$$=H_2[\delta(1-\varepsilon)+(1-\delta)\varepsilon]$$

$$H(Y|X)=-\sum_i\sum_j p(y_j|x_i)q(x_i)\log p(y_j|x_i)$$

$$=-\delta[(1-\varepsilon)\log(1-\varepsilon)+\varepsilon\log\varepsilon]-(1-\delta)[\varepsilon\log\varepsilon+(1-\varepsilon)\log(1-\varepsilon)]$$

$$=-(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon=H_2(\varepsilon)$$

则平均互信息量

$$I(X;Y)=H(Y)-H(Y|X)=H_2[\delta(1-\varepsilon)+(1-\delta)\varepsilon]-H_2(\varepsilon)$$

当信道固定，即ε 为恒值时，$I(X;Y)$是δ的∩函数，其曲线如图 2-11 所示。

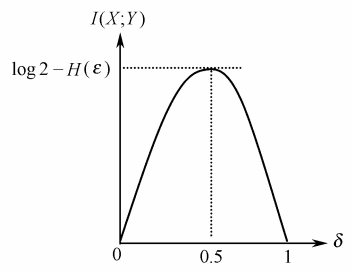


图 2-11　ε 为恒值时的 $I(X;Y)$曲线

当$\delta=0.5$ 时，$I(X;Y)$取得极大值，其值为 $\log 2-H_2(\varepsilon)$，这种情况对应等概分布，信源的平均不确定性最大。

当$\delta=0$ 或 1 时，这是确定信源的情况，通信得不到任何信息，即 $I(X;Y)=0$。

定理 2.2　当信源给定，即信源分布概率 $q(x)$固定时，平均互信息量 $I(X;Y)$是信道转移概率 $p(y|x)$的∪型凸函数。

根据式（2-24），在信源固定的情况下，如果给定两个信道转移概率 $p_1(y|x)$和 $p_2(y|x)$，它们分别对应平均互信息量 $I_1(X;Y)$和 $I_2(X;Y)$，记信道转移概率 $p(y|x)=\theta p_1(y|x)+(1-\theta)p_2(y|x)$（式中（$0\leqslant\theta\leqslant1$），对应平均互信息量 $I(X;Y)$，若 $I(X;Y)$是 $p(y|x)$的∪型凸函数，则应满足

$$I(X;Y)\leqslant\theta I_1(X;Y)+(1-\theta)I_2(X;Y) \qquad (2\text{-}40)$$

式（2-40）表示：均值的函数小于等于函数的均值，如图 2-12 所示。

证明：

$$I(X;Y)-\theta I_1(X;Y)+(1-\theta)I_2(X;Y)$$

$$=\sum_i\sum_j\left[\theta\cdot p_1(y_j|x_i)+(1-\theta)p_2(y_j|x_i)\right]\cdot q(x_i)\log\frac{\phi(x_i|y_j)}{q(x_i)}-$$

$$\theta\sum_i\sum_j p_1(x_iy_j)\log\frac{\phi_1(x_i|y_j)}{q(x_i)}-(1-\theta)\sum_i\sum_j p_2(x_iy_j)\log\frac{\phi_2(x_i|y_j)}{q(x_i)}$$

$$=\theta\sum_i\sum_j p_1(x_iy_j)\log\frac{\phi(x_i|y_j)}{\phi_1(x_i|y_j)}+(1-\theta)\sum_i\sum_j p_2(x_iy_j)\log\frac{\phi(x_i|y_j)}{\phi_2(x_i|y_j)}$$

$$\leqslant\theta\log\sum_i\sum_j p_1(x_iy_j)\frac{\phi(x_i|y_j)}{\phi_1(x_i|y_j)}+(1-\theta)\log\sum_i\sum_j p_2(x_iy_j)\frac{\phi(x_i|y_j)}{\phi_2(x_i|y_j)}$$

$$=\theta\log\sum_i\sum_j\omega_1(y_j)\phi(x_i|y_j)+(1-\theta)\log\sum_i\sum_j\omega_2(y_j)\phi(x_i|y_j)$$

$$=\theta\log\sum_j\omega_1(y_j)\sum_i\phi(x_i|y_j)+(1-\theta)\log\sum_j\omega_2(y_j)\sum_i\phi(x_i|y_j)$$

$$=\theta\log1+(1-\theta)\log1=0$$

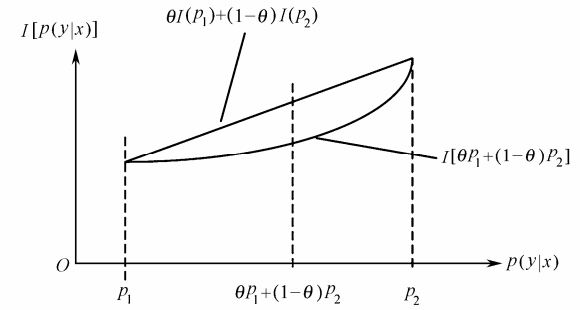


图 2-12　均值的函数小于等于函数的均值

所以

$$I(X;Y)\leqslant\theta I_1(X;Y)+(1-\theta)I_2(X;Y)$$　　证毕

定理 2.2 说明，信源固定以后，用不同的信道来传输同一信源符号时，在信道输出端获得的信息量是不同的。可见，对每一种信源一定存在一种最差的信道，此信道的干扰最大，而使输出端获得的信息量最小，因为∪型凸函数存在最小值。

【例 2.17】 在例 2.16 中，信源分布 $q(0)=\delta$，$q(1)=1-\delta$，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}$，已算出 $I(X;Y)=H_2[\delta(1-\varepsilon)+(1-\delta)\varepsilon]-H_2(\varepsilon)$。

当信源固定，即δ为恒值时，$I(X;Y)$是ε的∪型凸函数，其曲线如图 2-13 所示。

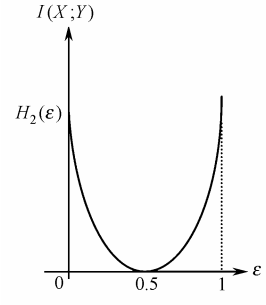


图 2-13　ε 为恒值时的 $I(X;Y)$曲线

由图可见，当$\varepsilon=0.5$ 时，$I(X;Y)$取得极小值，其值为 0。此时，由于信道干扰太大，以至于发送 0 或 1 时，接收端收到 0，1 的概率各为 0.5，这是一种最差的信道，信号完全被噪声淹没。

当$\varepsilon=0$ 时，这是无噪信道的情况，平均互信息量 $I(X;Y)$ 取得最大值，其值为 $H_2(\delta)$。

当$\varepsilon=1$ 时，这是强噪信道的情况，以至于发送 0 时信道收到 1，而发送 1 时反而收到 0，这时在判决时干脆将错就错，收到 0 判决为 1，收到 1 判决为 0，也能得到最大平均互信息量 $I(X;Y)=H_2(\delta)$。

2.4 *N*维扩展信源的熵和平均互信息量

2.4.1 *N*维扩展信源的熵

信源输出 N 维序列为 $\boldsymbol{x}=x_1\cdots\ x_i\ \cdots\ x_N$，序列中的每个符号 x_i（$i=1, 2, \cdots, N$）都是取值来自同一离散符号集的随机变量，$x_i\in\{a_0,a_1,\cdots,a_{k-1}\}$，记 $\boldsymbol{x}=x_1x_2\cdots x_N$ 的概率分布为 $q(\boldsymbol{x})$，则信源熵为

$$H(\boldsymbol{X})=\sum_{\boldsymbol{x}}q(\boldsymbol{x})\log\frac{1}{q(\boldsymbol{x})} \tag{2-41}$$

下面分两种情况来考虑。

1．信源离散无记忆

在信源输出符号离散无记忆的情况下，信源输出序列 $\boldsymbol{x}=x_1x_2\cdots x_N$ 的概率等于各随机变量 $x_i\in\{a_0,a_1,\cdots,a_{k-1}\}$（$i=1, 2, \cdots, N$）概率分布的连乘，即 $q(\boldsymbol{x})=\prod_{i-1}^{N}q(x_i)$，按式（2-41）可计算出该信源的熵为

$$\begin{aligned}
H(\boldsymbol{X})&=\sum_{\boldsymbol{x}}q(\boldsymbol{x})\log\frac{1}{\prod_{i=1}^{N}q(x_i)}\\
&=\sum_{i=1}^{N}\sum_{\boldsymbol{x}}q(\boldsymbol{x})\log\frac{1}{q(x_i)}\\
&=\sum_{i=1}^{N}\left(\sum_{x_1}\sum_{x_2}\cdots\sum_{x_i}\cdots\sum_{x_N}\right)\left(\prod_{i=1}^{N}q(x_i)\right)\log\frac{1}{q(x_i)}\\
&=\sum_{i=1}^{N}\left[\left(\sum_{x_1}q(x_1)\right)\left(\sum_{x_2}q(x_2)\right)\cdots\left(\sum_{x_i}q(x_i)\log\frac{1}{q(x_i)}\right)\cdots\left(\sum_{x_N}q(x_N)\right)\right]\\
&=\sum_{i=1}^{N}(1\cdot1\cdot\cdots\cdot H(x_i)\cdot\cdots\cdot1)\\
&=\sum_{i=1}^{N}H(x_i)
\end{aligned}$$

即

$$H(\boldsymbol{X})=\sum_{i=1}^{N}H(x_i) \tag{2-42}$$

在一般情况下，同一信源输出的各事件 x_1，x_2，…，x_N，其概率分布空间是相同的，即每一 x_i（$i=1, 2, \cdots, N$）取值为 $a_0, a_1, \cdots, a_{K-1}$ 的概率是相同的，所以由式（2-42）可得

$$H(\boldsymbol{X})\ =\ NH(X) \tag{2-43}$$

2．信源离散有记忆

在信源输出符号相互有关联的情况下，信源输出序列 $\boldsymbol{x}\ =\ x_1\ x_2\ \cdots\ x_N$ 的概率为 $p(\boldsymbol{x})\ =$

$p(x_1)p(x_2 \mid x_1)\,p(x_3 \mid x_1x_2)\ \cdots\ p(x_N \mid x_1x_2\cdots x_{N-1})$，相应地可以计算出其信源熵，即

$$H(\boldsymbol{X}) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1X_2) + \cdots + H(X_N \mid X_1X_2\cdots X_{N-1})$$

记为

$$H(\boldsymbol{X}) = \sum_{i=1}^{N} H\left(X_i \middle| \boldsymbol{X}_1^{i-1}\right) \tag{2-44}$$

这一结果称为熵的链规则。

根据熵的性质（6），条件熵小于等于无条件熵，即有

$$\begin{cases} H(X_2|X_1) \leq H(X_2) \\ H(X_3|X_1X_2) \leq H(X_3) \\ \vdots \\ H(X_N|X_1X_2\cdots X_{N-1}) \leq H(X_N) \end{cases} \tag{2-45}$$

将式（2-45）代入式（2-44），得

$$H(\boldsymbol{X}) \leq \sum_{i=1}^{N} H(X_i) \tag{2-46}$$

等号在信源无记忆（统计独立）时成立。

对于平稳信源，有 $p(x_2 \mid x_1) = p(x_3 \mid x_2) = \cdots = p(x_N \mid x_{N-1})$，即条件概率与时间起点无关，则式（2-44）可以写为

$$H(\boldsymbol{X}) = H(X_N) + H(X_N \mid X_{N-1}) + H(X_N \mid X_{N-1}X_{N-2}) + \cdots + H(X_N \mid X_{N-1}X_{N-2}\cdots X_1) \tag{2-47}$$

由于条件熵小于等于无条件熵，故下式成立

$$H(X_N) \geqslant H(X_N \mid X_{N-1}) \geqslant H(X_N \mid X_{N-1}X_{N-2}) \geqslant \cdots \geqslant H(X_N \mid X_{N-1}X_{N-2}\cdots X_1) \tag{2-48}$$

将式（2-48）代入式（2-47）得 $H(\boldsymbol{X}) \leqslant N H(X_N)$，不失一般性，记 $X_N = X$，即有

$$H(\boldsymbol{X}) \leqslant N H(X) \tag{2-49}$$

等号在信源无记忆时成立。

2.4.2 *N*维扩展信源的平均互信息量

先看二维情况：

$$\begin{aligned} I(X;Y_1Y_2) &= H(X) - H(X|Y_1Y_2) \\ &= H(X) + E\{\log p(x|y_1y_2)\} \\ &= H(X) + E\left\{\log \frac{p(xy_1y_2)}{p(y_1y_2)}\right\} \\ &= H(X) + E\left\{\log \frac{p(y_2|xy_1)\,p(x|y_1)\,p(y_1)}{p(y_2|y_1)\,p(y_1)}\right\} \\ &= H(X) + E\{\log p(x \mid y_1) - \log p(y_2 \mid y_1)\} + E\{\log p(y_2 \mid xy_1)\} \\ &= H(X) - H(X \mid Y_1) + H(Y_2 \mid Y_1) - H(Y_2 \mid XY_1) \\ &= I(X;Y_1) + I(X;Y_2 \mid Y_1) \end{aligned}$$

式中，符号 E 表示求统计均值。

由上面的计算结果得

$$I(X;Y_1Y_2) = I(X;Y_1) + I(X;Y_2 \mid Y_1) \tag{2-50}$$

类推到三维，则有

$$I(X;Y_1Y_2Y_3)=I(X;Y_1)+I(X;Y_2\mid Y_1)+I(X;Y_3\mid Y_1Y_2)$$

推广到 N 维矢量的情况，则有

$$I(X;\boldsymbol{Y})=I(X;Y_1)+I(X;Y_2\mid Y_1)+\cdots+I(X;Y_N\mid Y_1Y_2\cdots Y_{N-1})$$

记为

$$I(X;\boldsymbol{Y})=\sum_{i=1}^{N}I\left(X;Y_i\left|\boldsymbol{Y}_1^{i-1}\right.\right) \tag{2-51}$$

这一结果称为平均互信息量的链规则。

【例 2.18】 离散无记忆信源$\begin{bmatrix}X\\q(X)\end{bmatrix}=\begin{bmatrix}a_0 & a_1 & a_2\\1/2 & 1/4 & 1/4\end{bmatrix}$，求它的二次扩展信源的熵。

二次扩展信源的消息集为

$$\boldsymbol{x}=x^{(2)}=\{a_0a_0,\ a_0a_1,\ a_0a_2,\ a_1a_0,\ a_1a_1,\ a_1a_2,\ a_2a_0,\ a_2a_1,\ a_2a_2\}$$

记 $a_ia_j=b_k$，$i,j=0,1,2$，$k=0,1,\cdots,8$，则扩展信源可以记为

$$\begin{bmatrix}\boldsymbol{X}\\q(\boldsymbol{X})\end{bmatrix}=\begin{bmatrix}b_0 & b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8\\1/4 & 1/8 & 1/8 & 1/8 & 1/16 & 1/16 & 1/8 & 1/16 & 1/16\end{bmatrix}$$

可算得

$$H(\boldsymbol{X})=-\sum_{\boldsymbol{x}}q(\boldsymbol{x})\log q(\boldsymbol{x})=\frac{1}{4}\log4+4\times\frac{1}{8}\log8+4\times\frac{1}{16}\log16=3\ \text{（比特/符号）}$$

也可按下式计算

$$H(\boldsymbol{X})=\sum_{i=1}^{2}H(X_i)=2H(X)=2\left[\frac{1}{2}\log2+\frac{1}{4}\log4+\frac{1}{4}\log4\right]=3\ \text{（比特/符号）}$$

对于复合事件的平均互信息量，下式成立

$$\begin{cases}I(X;YZ)\geq I(X;Y)\\I(X;YZ)\geq I(X;Z)\end{cases} \tag{2-52}$$

式（2-52）说明复合事件的信息量大于单个事件的信息量。

证明：根据平均互信息量的链规则式（2-51），有

$$I(X;YZ)=I(X;Y)+I(X;Z\mid Y) \tag{2-53}$$

又根据平均互信息量的性质 1（平均互信息量的非负性），即

$$I(X;Z\mid Y)\geqslant0 \tag{2-54}$$

将式（2-54）代入式（2-53），即得 $I(X;YZ)\geqslant I(X;Y)$

类似地可以证得

$$I(X;YZ)\geqslant I(X;Z)$$

证毕

2.4.3 有关N维平均互信息量的两条定理

定理 2.3 如果信源离散无记忆，即信源输出序列 $\boldsymbol{x}=x_1x_2\cdots x_N$ 中各 x_i（$i=1,2,\cdots,N$）统计独立，则信道输入、输出符号序列间的平均互信息量 $I(\boldsymbol{X};\boldsymbol{Y})$大于等于各单个符号间平均互信息量的总和，即有

$$I(\boldsymbol{X};\boldsymbol{Y})\geq\sum_{i=1}^{N}I(X_i;Y_i) \tag{2-55}$$

证明：① 因为信源统计独立，根据式（2-42），有 $H(\boldsymbol{X})=\sum_{i=1}^{N}H(X_i)$；

② 由链规则式（2-44）得

$$H(\boldsymbol{X}|\boldsymbol{Y})=\sum_{i=1}^{N}H\left(X_i|\boldsymbol{X}_1^{i-1},\boldsymbol{Y}\right)\leq\sum_{i=1}^{N}H\left(X_i|Y_i\right) \tag{2-56}$$

所以

$$\begin{aligned}I(\boldsymbol{X};\boldsymbol{Y})&=H(\boldsymbol{X})-H(\boldsymbol{X}|\boldsymbol{Y})\\&=\sum_{i=1}^{N}H(X_i)-\sum_{i=1}^{N}H\left(X_i|\boldsymbol{X}_1^{i-1},\boldsymbol{Y}\right)\\&\geq\sum_{i=1}^{N}H(X_i)-\sum_{i=1}^{N}H\left(X_i|Y_i\right)\\&=\sum_{i=1}^{N}I\left(X_i;Y_i\right)\end{aligned}$$

证毕

定理 2.3 说明，在信道相互有关联的情况下，解码时不宜一个一个地解，而应一段一段地解（每段包括 N 个字符），这样得到的信息量大。

由上面的证明过程可看出，等号成立的条件为

$$H(\boldsymbol{X}|\boldsymbol{Y})=\sum_{i=1}^{N}H\left(X_i|Y_i\right)$$

这个条件要求信道也是离散无记忆的。

定理 2.4 若信道离散无记忆，则信道输入、输出符号序列间的平均互信息量 $I(\boldsymbol{X};\boldsymbol{Y})$小于等于各单个符号间平均互信息量的总和，即有

$$I(\boldsymbol{X};\boldsymbol{Y})\leq\sum_{i=1}^{N}I\left(X_i;Y_i\right) \tag{2-57}$$

证明：对于离散无记忆信道，有 $p(\boldsymbol{y}|\boldsymbol{x})=\prod_{i=1}^{N}p\left(y_i|x_i\right)$，因此

$$\begin{aligned}H(\boldsymbol{Y}|\boldsymbol{X})&=-\sum_{\boldsymbol{x},\boldsymbol{y}}p(\boldsymbol{xy})\log p(\boldsymbol{y}|\boldsymbol{x})\\&=-\sum_{\boldsymbol{x},\boldsymbol{y}}p(\boldsymbol{xy})\log\prod_{i=1}^{N}p\left(y_i|x_i\right)\\&=-\sum_{\boldsymbol{x},\boldsymbol{y}}p(\boldsymbol{xy})\log p\left(y_1|x_1\right)-\sum_{\boldsymbol{x},\boldsymbol{y}}p(\boldsymbol{xy})\log p\left(y_2|x_2\right)-\ldots-\sum_{\boldsymbol{x},\boldsymbol{y}}p(\boldsymbol{xy})\log p\left(y_N|x_N\right)\\&=-\sum_{x_1,y_1}p(x_1y_1)\log p\left(y_1|x_1\right)-\sum_{x_2,y_2}p(x_2y_2)\log p\left(y_2|x_2\right)-\ldots-\sum_{x_N,y_N}p(x_Ny_N)\log p\left(y_N|x_N\right)\\&=H\left(Y_1|X_1\right)+H\left(Y_2|X_2\right)+\ldots+H\left(Y_N|X_N\right)\\&=\sum_{i=1}^{N}H\left(Y_i|X_i\right)\end{aligned}$$

即有

$$H(\boldsymbol{Y}|\boldsymbol{X})=\sum_{i=1}^{N}H\left(Y_i|X_i\right) \tag{2-58}$$

另一方面，由于信源离散有记忆，由式（2-46)知 $H(\boldsymbol{X})\leq \sum_{i=1}^{N} H(X_i)$，这样的信源经过离散无记忆信道传输后，信道输出符号也是离散有记忆的，从而

$$H(\boldsymbol{Y})\leq \sum_{i=1}^{N} H(Y_i) \qquad (2\text{-}59)$$

将式（2-58）和式（2-59）代入平均互信息 $I(\boldsymbol{X};\boldsymbol{Y})$的表达式，有

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{Y}) - H(\boldsymbol{Y}\mid \boldsymbol{X}) \\ &\leq \sum_{i=1}^{N} H(Y_i) - \sum_{i=1}^{N} H(Y_i\mid X_i) \\ &= \sum_{i=1}^{N} I(X_i;Y_i) \end{aligned}$$

证毕

等号成立的条件为 $H(\boldsymbol{Y})=\sum_{i=1}^{N} H(Y_i)$，即要求信宿是离散无记忆的，这在离散无记忆信道传输条件下，说明信源分布是离散无记忆的。

本 章 小 结

1. 本章主要阐述了对各种信息量的度量，主要内容如下。

（1）信息量 $I(x_i) = -\log q(x_i)$

对自信息量求统计均值，得到熵 $H(X) = -\sum_i q(x_i)\log q(x_i)$。

（2）条件自信息量 $I(x_i \mid y_j) = -\log\phi(x_i \mid y_j)$。

对 $I(x_i \mid y_j)$求统计均值，得到条件熵 $H(X\mid Y) = -\sum_i\sum_j p(x_i y_j)\log\phi(x_i \mid y_j)$。

（3）条件自信息量 $I(y_j \mid x_i) = -\log p(y_j \mid x_i)$。

对 $I(y_j \mid x_i)$求统计均值，得到条件熵 $H(Y\mid X) = -\sum_i\sum_j p(x_i y_j)\log p(y_j\mid x_i)$。

（4）二维联合空间的自信息量 $I(x_i y_j) = -\log p(x_i y_j)$

对 $I(x_i y_j)$求统计均值，得到联合熵 $H(XY) = -\sum_i\sum_j p(x_i y_j)\log p(x_i y_j)$。

（5）两个事件之间的互信息 $I(x_i;y_j) = \log\dfrac{\phi(x_i\mid y_j)}{q(x_i)} = \log\dfrac{p(y_j\mid x_i)}{\omega(y_j)} = \log\dfrac{p(x_i y_j)}{q(x_i)\omega(y_j)}$

对 $I(x_i;y_j)$求统计均值，得到平均互信息量

$$I(X;Y) = H(X) - H(X\mid Y) = H(Y) - H(Y\mid X) = H(X) + H(Y) - H(XY)$$

2. 前面针对个别事件的各种信息量，可看做过渡物理量，我们关心的还是它们所对应的统计平均值，上面列出的 5 个有关信息量的统计平均值，可用维拉图帮助理清它们之间的关系。

3. 本章介绍了三个重要定理，请予关注。

（1）极大熵定理;

（2）定理 2.1；信道给定，$I(X;Y)$是信源的∩型凸函数;

（3）定理 2.1：信源给定，$I(X;Y)$是信道的∪型凸函数。

思考题与习题

2.1 信源在何种分布时，熵值最大？又在何种分布时，熵值最小？

2.2 平均互信息量 $I(X;Y)$与信源分布 $q(x)$有何关系？与 $p(y|x)$又是什么关系？

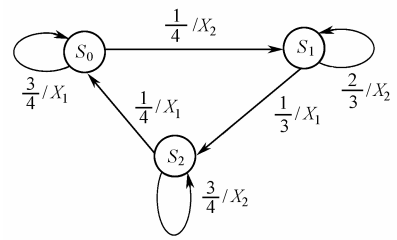
2.3 熵是对信源什么物理量的度量？

2.4 设信道输入符号集为$\{x_1, x_2, \cdots, x_k\}$，则平均每个信道输入符号所能携带的最大信息量为多少？

2.5 根据平均互信息量的链规则，写出 $I(X; YZ)$的表达式。

2.6 互信息量 $I(x; y)$有时候取负值，是由于信道存在干扰或噪声的原因，这种说法对吗？

2.7 一个马尔可夫信源如题图 2-1 所示，求稳态下各状态的概率分布和信源熵。



题图 2-1 马尔可夫信源

2.8 一个马尔可夫信源，已知：

$$p(x_1|x_1)=\frac{2}{3},\ p(x_2|x_1)=\frac{1}{3},\ p(x_1|x_2)=1,\ p(x_2|x_2)=0$$

试画出它的香农线图，并求出信源熵。

2.9（1）对于离散无记忆信源 DMS $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ p & 1-p \end{bmatrix}$，试证明：

$$H(X)=H_2(p)=-p\log p-(1-p)\log(1-p)$$

当 $p=\frac{1}{2}$时，$H(X)$达到最大值。

（2）对（1）中的 DMS，考虑它的二次扩展信源 $X^{(2)}=\{(x_1x_1),(x_1x_2),(x_2x_1),(x_2x_2)\}$，证明：

$$H(X^{(2)})=2H(X)$$

2.10 一副扑克牌（不用大小王），试问：

（1）任意特定排列给出的信息量是多少？

（2）从 52 张牌中抽取 13 张，所给出的点数都不相同时得到多少信息量？

（3）从 52 张牌中任意抽取 1 张，然后放回，结果视为从 DMS 中取得样本，这个 DMS 的熵为多少？

（4）若（3）中不计颜色，熵又为多少？

2.11（1）一个无偏骰子，掷骰子的熵为多少？

（2）如果骰子被改造使得某点出现的概率与其点数成正比，那么熵为多少？

（3）一对无偏骰子，各掷一次，得到总点数为 7，问得到多少信息量？

2.12 一个盒子中放有 100 个球，其中 60 个球是黑色的，40 个球是白色的。

（1）随机摸取一球，求获得的自信息量；

（2）进行放回摸取 n 次，求这 n 次所得到的平均自信息量。

2.13 已知平均每 100 人中有 2 人患有某种病，为了查明病情进行某项指标的化验。化验结果对于病人总是阳性的，而对于健康人来说，这项指标有一半可能为阳性，一半可能为阴性。

问这项化验对于查明病情提供了多少信息?

2.14 一个 8 元编码系统，码长为 4，每个码字的第一个字符相同（用于同步），若每秒产生 1000 个码字，求信息传输速率 R_t。

2.15 一副拼板，其中有 3 块圆形，4 块方形，5 块三角形，随机排成一行，每一种排列都是等可能的，如果要求不能有 2 块方形相邻，可以得到多少关于拼板排列的信息?

2.16 设有一个传输系统，等概传输 0，1，2，3，4，5 六个数字，奇数在传输时以 0.5 的概率错成其他奇数，偶数能正确接收，求此传输系统的平均互信息量。

2.17 等概信源消息集 $u_0, u_1, \cdots, u_7$，编码为 u_0=000, u_1=001, …, u_7=111，通过错误概率为 p 的二进制对称信道 BSC 传输，在接收 u_4=100 的过程中，求

（1）1 与 u_4 之间的互信息量;

（2）10 与 u_4 之间的互信息量;

（3）100 与 u_4 之间的互信息量。

2.18 离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1=0 & x_2=1 & x_3=2 \\ 2/9 & 3/9 & 4/9 \end{bmatrix}$，输出符号串 $\boldsymbol{x}$ = 02101220102021002120 2，试求:

（1）信源熵 H(X);

（2）自信息量 $I(\boldsymbol{x})$;

（3）符号串 $\boldsymbol{x}$ 中平均每个符号携带的信息量。

2.19 给定信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.6 & 0.4 \end{bmatrix}$,

（1）该信源是平稳信源吗？计算信源熵;

（2）计算 H(X^3)，并列出信源 $\begin{bmatrix} X^3 \\ q(X^3) \end{bmatrix}$;

（3）计算 H($X_3 \mid X_1X_2$)及 N 维扩展信源在 N 趋于无穷时的熵 $\lim\limits_{N\to\infty} H_N(\boldsymbol{X})$。

2.20 求出概率分布$\{q_k, k=0,1,2,\cdots\}$，使在限制条件 $\sum\limits_{k=0}^{\infty} k \cdot q_k = A$ 下，熵达到最大值，并计算该最大值。

2.21 X,Y,Z 为概率空间，证明下述关系式成立，并给出等号成立的条件。

（1）$H(YZ \mid X) \leqslant H(Y \mid X) + H(Z \mid X)$

（2）$H(YZ \mid X) = H(Y \mid X) + H(Z \mid XY)$

（3）$H(X \mid Z) \leqslant H(X \mid Y) + H(Y \mid Z)$

2.22 对任意概率事件集 X,Y,Z，证明下述三角不等式成立:

$$\frac{H(X|Y)}{H(XY)} + \frac{H(Y|Z)}{H(YZ)} \geq \frac{H(X|Z)}{H(XZ)}$$

2.23 令 $X \to Y \to Z$ 为马尔可夫链，证明:

（1）$I(X; Z \mid Y) = 0$

（2）$I(XY;Z) = I(Y;Z)$

（3）$I(Y; Z \mid X) = I(Y; Z) - I(X; Z)$

（4）$I(Y; Z \mid X) \leqslant I(Y; Z)$

2.24　X, Y, Z 为离散随机变量，它们之间有如下关系：$Z=X+Y$，其中 X 和 Y 统计独立，证明:

（1）$H(X)\leqslant H(Z)$，等式仅当 Y 为常量时成立;

（2）$H(Y)\leqslant H(Z)$，等式仅当 X 为常量时成立;

（3）$H(Z)\leqslant H(XY)\leqslant H(X)+H(Y)$，等式仅当 X, Y 中任意一个为常量时成立;

（4）$I(X;Z)=H(Z)-H(Y)$;

（5）$I(XY;Z)=H(Z)$。

2.25　X, Y, Z 为二元随机变量，它们之间有如下关系：$Z=XY$，其中 X 和 Y 统计独立且各自等概分布，计算:

（1）$H(Z)$，$H(XY)$，$H(XYZ)$;

（2）$H(X\mid Y)$，$H(X\mid Z)$，$H(Z\mid Y)$;

（3）$I(X;Z)$，$I(X;Y\mid Z)$，$I(Z;X\mid Y)$，$I(XY;Z)$。

2.26　X, Y, Z 为二元随机变量，它们之间有如下关系：$Z=X\oplus Y$（为模二运算），其中 X 和 Y 统计独立且各自等概分布，计算:

（1）$H(Z)$，$H(XY)$，$H(XYZ)$;

（2）$H(X\mid Y)$，$H(X\mid Z)$，$H(Z\mid Y)$;

（3）$I(X;Z)$，$I(X;Y\mid Z)$，$I(Z;X\mid Y)$，$I(XY;Z)$。

2.27　证明几何分布 $\begin{bmatrix} x \\ q(x) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots \\ p & p(1-p) & \cdots & p(1-p)^{i-1} & \cdots \end{bmatrix}$ 的熵为 $H(X)=\dfrac{H_2(p)}{p}$。

2.28　应用熵与互信息的链规则证明：$I(\boldsymbol{X};\boldsymbol{Y})=\sum_{i=1}^{n}\sum_{j=1}^{n}I\left(x_i;y_j\left|\boldsymbol{X}_1^{i-1}\boldsymbol{Y}_1^{j-1}\right.\right)$。

2.29　信源消息集 $X=\{0,1\}$，信宿消息集 $Y=\{0,1\}$，信源等概分布，通过二进制信道 $\left[p(y|x)\right]=\begin{bmatrix} 0.76 & 0.24 \\ 0.32 & 0.68 \end{bmatrix}$ 传输，求:

（1）该系统的平均互信息量;

（2）接收到 $y=0$ 后，所提供的关于 x 的平均互信息量 $I\{X;0\}$。

2.30　一传输系统的输入符号集 $X=\{x_0, x_1, x_2, x_3\}$，输出符号集 $Y=\{y_0, y_1, y_2\}$，输入符号与输出符号的联合概率 $p(xy)$ 用下述矩阵表示:

$$\left[p(xy)\right]=\begin{array}{c} \\ x_0 \\ x_1 \\ x_2 \\ x_3 \end{array}\begin{array}{c} \begin{array}{ccc} y_0 & y_1 & y_2 \end{array} \\ \begin{bmatrix} 0.1 & 0 & 0 \\ 0.2 & 0.1 & 0 \\ 0 & 0.3 & 0.2 \\ 0 & 0 & 0.1 \end{bmatrix} \end{array}$$

计算 $H(X)$，$H(Y)$，$H(Y\mid X)$，$H(XY)$ 及 $I(X;Y)$，并与维拉图对照。

第 3 章
离散信源无失真编码

内容提要

用尽可能少的符号来传输信源消息，目的是提高传输效率，这是信源编码应考虑的问题，本章讨论在不允许失真情况下的信源编码。等长编码定理给出了在等长编码条件下，其码长的下限值，变长编码定理（香农第一定理）给出了信源无失真变长编码时其码长的上、下限值。本章还介绍了三种通用信源编码方法：香农编码法、Fano 编码法和霍夫曼编码法。

知识要点

信息传输率、克拉夫特不等式、等长编码定理、变长编码定理、编码效率、无失真编码方法。

教学建议

本章介绍了两个重要定理：等长编码定理和变长编码定理，实际上就是在满足无失真和码长尽可能短这两个条件下，给出了码长的上、下限值，定理的证明冗长烦琐，所以不予证明，只讲述定理的物理意义。通过例题掌握三种无失真信源编码方法，用编码效率衡量其优劣。建议学时数为 6 学时。

3.1 概　　述

各种通信系统，尽管它们的形式和用途各不相同，但都可以归结为图 1-1 所示的模型。其中信源、信宿和信道是事先给定的，信源是产生信息的源泉，信宿用于接收信息，而信道则用于传输信息。

通信要完成的任务就是，将信源的输出经由信道在接收端精确地或近似地重现出来，当然，这种传输最好是高效的、可靠的。为了实现高质量、高效率的通信，引入了信源编码和信道编码，这些都是由人来设计完成的，通信质量的好坏，很大程度上取决于编码、译码过程设计的好坏。信源编码和信道编码主要需要解决以下两个问题。

1．提高传输效率

用尽可能少的信道传输符号来传递信源消息，目的是提高传输效率，这是信源编码应考虑的主要问题。这里又分两种情况讨论，即允许接收信号有一定的失真或不允许失真。

在讨论信源编码时，暂且不考虑信道的干扰，如果不允许信号失真，即要求将信源输出在接收端精确地重现出来，这就要对信源进行无失真编码，保证信源输出所携带的信息全部无损地送达信宿，无失真编码只是对信源的冗余度进行压缩，并不会改变信源的熵，它能保证码元序列经无扰信道传输后得到无失真恢复，这就是本章要讨论的内容。

在许多实际情况中，信宿方并不要求完全精确地复现信源输出的原信号，例如，在电话通信中，只要将通话内容送达对方就可以了，对音质并没有太高的要求。事实上，在信道存在干扰的情况下，要完全精确地复现信源输出也几乎是做不到的，在这种情况下，允许接收信号有一定的失真。为提高传输效率，我们就可以事先对信源进行压缩编码，压缩到的什么程度由允许失真的程度来确定，有失真编码压缩了信源的熵，这是第 6 章要讨论的问题。

总之，信源编码以提高传输效率为主。

2．增强通信的可靠性

信号在信道的传播过程中总不可避免地会受到各种干扰，在这种情况下，如何增加信号的抗干扰能力，提高传输的可靠性，是信道编码主要考虑的问题。解决这一问题，一般采用冗余编码法，即按照一定的编码规则事先给信码加上一定的冗余度（检测位），赋予信码自身一定的纠错和检错能力，只要采取适当的信道编码和译码措施，就可使信道传输的差错率降到允许的范围之内，这是后面章节要讨论的问题。

综上所述，提高抗干扰能力往往是以降低信息传输效率为代价的，而为了提高传输效率又往往削弱了其抗干扰能力。这样，设计者在取舍之间就要进行均衡考虑，当然，香农已经在理论上证明，至少存在着某种最佳编码方法，可以有效地解决上述矛盾。

信源编码实际上是对信源的原始符号按一定的数学规则进行的一种代码变换。

信源编码包括以下两个功能：

① 将信源符号变换成适合信道传输的符号；

② 压缩信源冗余度，提高传输效率。

一般来说，信源编码可归纳为如图 3-1 所示的模型。

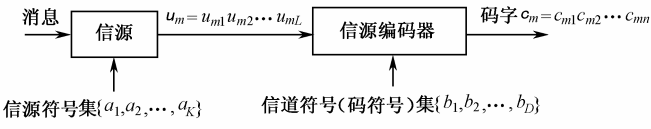


图 3-1　信源编码器模型

$\{a_1, a_2, \cdots, a_K\}$为信源符号集，信源消息用信源输出的符号序列 $\boldsymbol{u}_m = u_{m1}\, u_{m2} \cdots\, u_{mL}$ 表示，L 是信源序列长度，序列中每一个符号 u_{ml} 都是取自信源符号集，即 $u_{ml} \in \{a_1, a_2, \cdots, a_K\}$，$l = 1, 2, \cdots, L$，$m = 1, 2, \cdots, M$，$M$ 为信源消息个数。

$\{b_1，b_2，\cdots，b_D\}$是适合信道传输的 D 个符号，用做信源编码器的编码符号。信源编码器按一定的数学规则对 $\boldsymbol{u}_m = u_{m1}\, u_{m2} \cdots\, u_{mL}$ 进行编码，输出码字 $\boldsymbol{c}_m = c_{m1}\, c_{m2} \cdots\, c_{mn}, c_{mk} \in \{b_1, b_2, \cdots, b_D\}$，$k = 1, 2, \cdots, n$，$n$ 表示码字长度，简称码长。

从数学角度看，信源编码相当于一个 $\boldsymbol{u} \to \boldsymbol{c}$ 的映射，这种映射可以是一一对应的（无压缩编码），也可以是多对一的（压缩编码）。

【例 3.1】　天气预报：信源有 4 个消息{晴，阴，雨，雪 }待发，信源符号集为$\{a_1，a_2，a_3，a_4\}$，因为消息个数与信源符号集个数相等，故可以用单符号来表示消息，有 $\begin{Bmatrix} 晴 & 阴 & 雨 & 雪 \\ a_1 & a_2 & a_3 & a_4 \end{Bmatrix}$，如果采用二进制信道，即码符号集为{0,1}，则可将这 4 个消息变换成 4 个 2 位的二进制代码，即

a_1：00　　a_2：01　　a_3：10　　a_4：11

{00, 01, 10, 11}是分别代表$\{a_1，a_2，a_3，a_4\}$这 4 个消息的码字，码长为 2。

在此例中，{晴，阴，雨，雪 }是信源消息，$\{a_1，a_2，a_3，a_4\}$是信源符号，码符号集{0,1}是信源编码器的输出符号，即信道的输入符号，信源编码就是从$\{a_1，a_2，a_3，a_4\}$到{0,1}的一种映射，由于码符号集仅含{0,1}二个数字，故相应得到的码字称为二进制码字。

【例 3.2】　汉字电报是把常用的 10 000 个汉字变成 10 000 个 4 位的十进制数，再把每一个十进制数字 0,1,…,9 变换成一个 5 位的二进制等重代码组，即

$$\begin{array}{ccccccccc} u_1 & \Rightarrow & 0000 & \Rightarrow & 01101 & 01101 & 01101 & 01101 \\ u_2 & \Rightarrow & 0001 & \Rightarrow & 01101 & 01101 & 01101 & 01011 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{10000} & \Rightarrow & 9999 & \Rightarrow & 10011 & 10011 & 10011 & 10011 \end{array}$$

在此例中，10 000 个汉字（用 $u_1, u_2, \cdots, u_{10000}$ 表示）是信源消息，{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}是信源符号集，{0000，0001，…，9999}则是对这 10 000 个汉字的标识，码符号集为{0,1}，从{0000，0001，…，9999}到{0,1}的映射称为信源编码，得到的也是二进制编码。

3.1.1　码的分类

如前所述，信源编码可看成是从信源符号集到码符号集的一种映射，即将信源符号集中的每个元素（可以是单符号，也可以是符号序列）映射成一个长度为 n 的码字。

对于同一个信源，编码方法是多样的。

【例 3.3】　用$\{u_1，u_2，u_3，u_4\}$表示信源的 4 个消息，码符号集为{0,1}，表 3-1 列出了该信源的几种不同编码。

表 3-1　同一信源的几种不同编码

信 源 消 息	各消息概率	码 1	码 2	码 3	码 4
u_1	$q(u_1)$	00	00	0	1
u_2	$q(u_2)$	11	01	1	10
u_3	$q(u_3)$	10	10	00	100
u_4	$q(u_4)$	11	11	11	1000

一般地，可以将码简单地分成如下几类。

1. 二元码

若码符号集为{0，1}，则码字就是二元序列，称为二元码，二元码通过二进制信道传输，是数字通信和计算机通信中最常见的一种码，表 3-1 列出的 4 种码都是二元码。

2. 等长码

在一组码字集合 $\boldsymbol{C}$ 中的所有码字 $\boldsymbol{c}_m$（$m = 1, 2, \cdots, M$），其码长都相同，则称这组码 $\boldsymbol{C}$ 为等长码，表 3-1 中列出的码 1、码 2 就是码长 $n = 2$ 的等长码。

3. 变长码

若码字集合 $\boldsymbol{C}$ 中的所有码字 $\boldsymbol{c}_m$（$m = 1, 2, \cdots, M$），其码长不都相同，则称码 $\boldsymbol{C}$ 为变长码，表 3-1 中列出的码 3、码 4 就是变长码。

4. 奇异码

对奇异码来说，从信源消息到码字的影射不是一一对应的，即在一组码 $\boldsymbol{C}$ 中，存在着不同的信源消息用同一码字来表示的情况。例如表 3-1 中的码 1，信源消息 u_2 和 u_4 都用码字 11 对其编码，因此这种码就是奇异码，奇异码不具备唯一可译性。

5. 非奇异码

从信源消息到码字的影射是一一对应的，每一个不同的信源消息都用不同的码字对其编码，例如，表 3-1 中的码 2、码 3 和码 4 都是非奇异码。

6. 原码C的N次扩展码

原码 $\boldsymbol{C}$ 的 N 次扩展码中的每个元素，是 N 次扩展信源中的序列所对应的 N 个码字组成的序列。原码的编码示意图如图 3-1 所示，图 3-2 是 N 次扩展码的编码示意图。

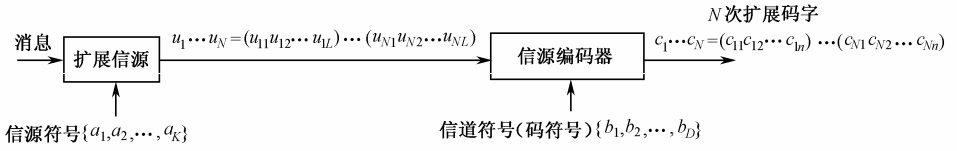


图 3-2　N 次扩展信源编码器模型

原码 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2,\ldots, \boldsymbol{c}_M\}$：信源有 M 个消息，用信源符号序列 $\boldsymbol{u}_m = u_{m1}\, u_{m2} \ldots u_{mL}$ 表示，$m = 1, 2, \cdots, M$，L 是信源输出序列长度，信源编码器将 $\boldsymbol{u}_m$ 编码为 $\boldsymbol{c}_m = c_{m1}\, c_{m2} \ldots c_{mn}$，$n$ 为码字长度。

原码的 N 次扩展码是将信源进行 N 次扩展得到的新信源符号序列 $\boldsymbol{u}^{(N)}=\boldsymbol{u}_1\cdots\boldsymbol{u}_N=(u_{11}u_{12}\cdots u_{1L})\cdots(u_{N1}u_{N2}\cdots u_{NL})$（每个 $\boldsymbol{u}^{(N)}$ 都是原码信源输出的 N 个序列的级联），对应码符号序列 $\boldsymbol{c}^{(N)}=\boldsymbol{c}_1\cdots\boldsymbol{c}_N=(c_{11}c_{12}\cdots c_{1n})\cdots(c_{N1}c_{N2}\cdots c_{Nn})$（每个 $\boldsymbol{c}^{(N)}$ 都是原码 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M\}$ 中的 N 个元素的级联），记集合 $\boldsymbol{C}^{(N)}=\{\boldsymbol{c}_1^{(N)},\boldsymbol{c}_2^{(N)},\cdots\}$，$\boldsymbol{C}^{(N)}$ 即原码 $\boldsymbol{C}$ 的 N 次扩展码。

7. 唯一可译码

定义 3.1 如果码的任意 N 次扩展码都是非奇异码，则称该码为唯一可译码。

表 3-1 中的码 1 显然不是唯一可译码，因为信源消息 u_2 和 u_4 对应同一个码字。

对于定长码，若原码是唯一可译码，则它的 N 次扩展码也是唯一可译的，而对于变长码则不尽然，见表 3-2。

表 3-2 同一信源的几种不同变长编码

信源消息	各消息概率	码 1	码 2	码 3
u_1	$q(u_1)$	0	1	1
u_2	$q(u_2)$	1	10	01
u_3	$q(u_3)$	00	100	001
u_4	$q(u_4)$	11	1000	0001

对于表 3-2 中的码 1，由于它的每一个信源符号对应不同的码字，所以它本身唯一可译，但将它进行二次扩展后得到的二次扩展码就不唯一可译，例如，二次扩展码中的 u_1u_3 和 u_3u_1 对应同一个码字 000，u_2u_4 和 u_4u_2 对应同一个码字 111，因此码 2 也不是唯一可译码。

表 3-2 中的码 2、码 3 不仅本身是唯一可译码，进行 N 次扩展后得到的 N 次扩展码也是唯一可译的，按照定义 3.1，码 2、码 3 是唯一可译码。

8. 即时码

对于变长码，有如下定义。

定义 3.2 对于码字 $\boldsymbol{c}=c_1c_2\cdots c_n$，称 $\boldsymbol{c}=c_1c_2\cdots c_i\,(i<n)$ 为码字 $\boldsymbol{c}$ 的字头（前缀）。

在表 3-2 的码 3 中，1 是 10 的前缀，10 是 100 的前缀，100 是 1000 的前缀。

定义 3.3 若码中任一码字都不是另一码字的字头，称该码为异字头码（无前缀码）。

表 3-3 中的码 3 是无前缀码，而其余的码 1 和码 2 都不是无前缀码。

对于码 3，收到“1”后就知道一个码字已经完结，马上就能判断出是什么码字，无须等待下一个符号到达，所以无前缀码能够即时译码，称为即时可译码，简称即时码。

而对于码 2，收到“1”后，并不能立即做出判决，就是收到“10”也不能立即做出判决，因为若下一次收到“1”，则可判断前面发出的是 a_2；若下一个是“0”，则还要收到下面的码元才能做出判决。所以非异字头码不能即时译码，称为非即时码，由于非异字头码中的一些码字是另一些码字的延长，故也称延长码。

显然，即时码是唯一可译码，而唯一可译码不一定是即时码。

即时码可用树图法来构造。

对于 D 进制码，从树根出发，可引出 D 根树枝，分别赋予每根树枝一个不同的码符号，树枝的端点为节点，每一节点又可引出 D 根分枝，又分别赋予这 D 根分枝中的每根一个不同的码符号，如某一节点被定为码字后，它就不再引出树枝，该节点称为终节点。码字就是从树

根出发到达终节点所对应的码符号序列。

【例 3.4】　用树图法表示表 3-2 中的码 3，如图 3-3 所示（$D=2$）。

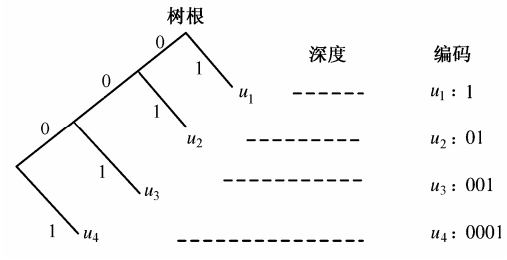


图 3-3　用树图法编码

【例 3.5】　用树图法表示表 3-3 中的码字，如图 3-4 所示。

表 3-3　信源编码

信 源 符 号	码　　字
u_1	00
u_2	01
u_3	100
u_4	101
u_5	110
u_6	111

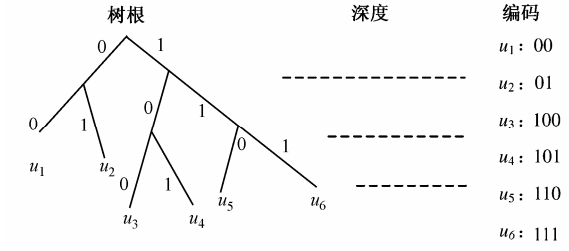


图 3-4　用树图法编码

图 3-5 是码的分类结构图。

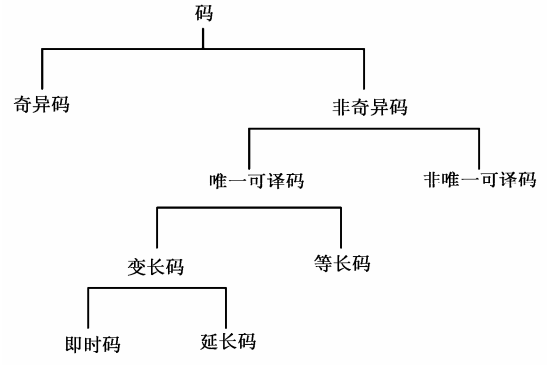


图 3-5　码的分类结构图

由上面的结构图可看出，码可分为奇异码和非奇异码两大类，我们只讨论非奇异码。非奇异码又分为唯一可译码和非唯一可译码两大类，我们只讨论唯一可译码。

3.1.2　平均码长的计算

对于变长码，码集 $\boldsymbol{C}$ 的平均码长用符号 $\bar{n}$ 表示，定义码 $\boldsymbol{C}$ 中每个码字为 $\boldsymbol{c}_m$（$m=1, 2, \cdots, M$），其码长的概率加权平均值为

$$\overline{n} \triangleq \sum_{m=1}^{M} n_m p(\boldsymbol{c}_m) \tag{3-1}$$

式中，n_m 是码字 $\boldsymbol{c}_m$ 所对应的码字的长度，$p(\boldsymbol{c}_m)$ 是码字 $\boldsymbol{c}_m$ 出现的概率。

对于等长码，由于码集 $\boldsymbol{C}$ 中的每个码字的码长都相同，平均码长就等于每个码字的码长

$$\overline{n}_1 = \overline{n}_2 = 2\times(0.4+0.2+0.2+0.2) = 2$$

在例 3.3 中，若给定各消息概率 $q(u_1)$= 0.4，$q(u_2)$= 0.2，$q(u_3)$= 0.2，$q(u_4)$= 0.2，则可算出码 1 和码 2 的平均码长 $\overline{n}_1 = \overline{n}_2 = 2\times(0.4+0.2+0.2+0.2) = 2$

码 3 的平均码长 $\overline{n}_3 = \sum_{m=1}^{4} n_m q(u_m) = 1\times0.4+1\times0.2+2\times0.2+2\times0.2 = 1.4$

码 4 的平均码长 $\overline{n}_4 = \sum_{m=1}^{4} mq(u_m) = 1\times0.4+2\times0.2+3\times0.2+4\times0.2 = 2.2$

N 次扩展码的平均码长 $\overline{n}$ 等于扩展码中码字长度的概率加权平均值。

对于 2 次扩展码，有

$$\overline{n} \triangleq \sum_{m}\sum_{s}(n_m+n_s)\,q(u_m)q(u_s) \tag{3-2}$$

设 n_m, n_s 分别是原信源消息 u_m, u_s 所对应的码长，c_m, c_s 是 u_m, u_s 所对应的码字，则式（3-2）中的 n_m+n_s 是扩展后新的信源序列 u_mu_s 所对应的码字 c_mc_s 的长度，$q(u_m)\,q(u_s)$ 是 c_mc_s 出现的概率。

【例 3.6】 考虑表 3-2 中码 3 的 2 次扩展码，给定消息概率 $q(u_1)=0.4$，$q(u_2)=0.2$，$q(u_3)=0.2$，$q(u_4)=0.2$，其二次扩展信源、二次扩展码及概率分布见表 3-4。

表 3-4　二次扩展码

二次扩展信源	二次扩展码	概 率 分 布	码　　长
u_1u_1	11	0.4×0.4=0.16	2
u_1u_2	101	0.4×0.2=0.08	3
u_1u_3	1001	0.4×0.2=0.08	4
u_1u_4	10001	0.4×0.2=0.08	5
u_2u_1	011	0.2×0.4=0.08	3
u_2u_2	0101	0.2×0.2=0.04	4
u_2u_3	01001	0.2×0.2=0.04	5
u_2u_4	010001	0.2×0.2=0.04	6
u_3u_1	0011	0.2×0.4=0.08	4
u_3u_2	00101	0.2×0.2=0.04	5
u_3u_3	001001	0.2×0.2=0.04	6
u_3u_4	0010001	0.2×0.2=0.04	7
u_4u_1	00011	0.2×0.4=0.08	5
u_4u_2	000101	0.2×0.2=0.04	6
u_4u_3	0001001	0.2×0.2=0.04	7
u_4u_4	00010001	0.2×0.2=0.04	8

计算出扩展码的平均码长为

$$\bar{n} \triangleq \sum_m \sum_s (n_m + n_s)\, q(u_m) q(u_s)$$

$$= 2\times 0.16+(3+4+5+3+4+5)\times 0.08 +(4+5+6+5+6+7+6+7+8) \times 0.04$$

$$= 4.4$$

3.1.3 信息传输速率

信道的信息传输速率为信道单位时间内所传输的实际信息量。

对于非压缩码（无失真编码），信息传输速率实际上就是信源的熵率（时间熵）。

若信息量以比特为单位，时间以秒为单位，则信息传输率定义为

$$R_t = \frac{H(X)}{t\bar{n}} \text{（比特/秒）} \tag{3-3}$$

式中，$H(X)$为信源熵；$\bar{n}$ 为编码后的平均码长；t 为传输一个码符号的时间。

若信息量以比特为单位，时间以码元时间（传输一个码符号的时间）为单位，则信息传输率记为

$$R_D = \frac{H(X)}{\bar{n}} \text{（比特/码元时间）} \tag{3-4}$$

在信息论中常用 R_D 作为信息传输率，而在其他场合则用 R_t 较多。

【例 3.7】 不计标点符号，英语由 26 个字母及 1 个字母间隔组成，据统计，这 27 个字符出现的概率见表 3-5。若每个符号分别用 5 位二进制代码表示，“0”和“1”都占一码元时间，用二进制信道传输，求信息传输率 R_D。

表 3-5 英文字母出现的概率

序	符 号	出 现 概 率	序	符 号	出 现 概 率
1	空格	0.181 7	15	M	0.021 05
2	E	0.107 3	16	U	0.020 10
3	T	0.085 6	17	G	0.016 33
4	A	0.066 8	18	Y	0.016 23
5	O	0.065 4	19	P	0.016 23
6	N	0.058 1	20	W	0.012 60
7	R	0.055 9	21	B	0.011 79
8	M	0.051 0	22	V	0.007 52
9	S	0.049 9	23	K	0.003 44
10	H	0.043 05	24	X	0.001 36
11	D	0.031 00	25	J	0.001 08
12	L	0.027 25	26	Q	0.000 99
13	F	0.023 95	27	Z	0.000 63
14	C	0.022 60			

这是一种无失真编码

$$R_D = \frac{H(X)}{n} = \frac{-\sum_{i=1}^{27} q(i)\log q(i)}{n} = \frac{4.03}{5} = 0.8 \text{（比特/码元时间）}$$

【例 3.8】　给定信源$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 1/4 & 1/4 & 1/8 & 1/8 & 1/16 & 1/16 & 1/16 & 1/16 \end{bmatrix}$，为提高传输效率，使平均码长尽可能短，遵照概率大取码长短，概率小取码长长的原则对上述信源进行二进制不等长编码，得到$\begin{cases} x_0:00 & x_4:1000 \\ x_1:01 & x_5:1001 \\ x_2:110 & x_6:1110 \\ x_3:101 & x_7:1111 \end{cases}$，上述码字用二进制信道传输，求信息传输率 R_D。

$$H(X)=-\left(2\times\frac{1}{4}\log\frac{1}{4}+2\times\frac{1}{8}\log\frac{1}{8}+4\times\frac{1}{16}\log\frac{1}{16}\right)=2.75\log 2=2.75\ \text{（比特/符号）}$$

$$\bar{n}=2\times 2\times\frac{1}{4}+2\times 3\times\frac{1}{8}+4\times 4\times\frac{1}{16}=2.75\ \text{（码元/符号）}$$

$$R_D=\frac{H(X)}{\bar{n}}=\frac{2.75}{2.75}=1\ \text{（比特/码元时间）}$$

由于每个二进制码符号所能携带的最大信息量为 1 比特（等概情况下），所以上面是一种最佳编码（紧致码）。紧致码是一种理想编码，只有在信源概率具有某种特殊分布时才有可能得到这种理想编码。在上面给出的信源中，可以看出，信源中每个消息的概率都可以表示为 2^{-h} 的形式，其中 h 为整数。

3.2　等长码及等长编码定理

对信源实现无差错编码，要求信源符号与码字一一对应。

下面考虑对一简单信源 S 进行等长编码，设信源符号集有 K 个符号，码符号集含 D 个符号，码字长度记为 n，则这 D 个码符号可组成 D^n 个码字。要对单符号信源 S 进行等长无差错编码，得到的码集 $\boldsymbol{C}$ 应是唯一可译的，而要得到唯一可译码，必须满足下式

$$K\leqslant D^n$$

考察表 3-1 中的码 2，看是否满足 $K\leqslant D^n$，用信源符号$\{a_1, a_2, a_3, a_4\}$分别表示信源的 4 个消息$\{u_1, u_2, u_3, u_4\}$，码符号集为$\{0, 1\}$，对于此码，有 $K=4$，$D=2$，码长 $n=2$，满足 $K\leqslant D^n$。

当然 $K\leqslant D^n$ 只是唯一可译码的必要条件，而不是充分条件，例如，表 3-1 中的码 1 也满足 $K\leqslant D^n$，但它不是唯一可译码。

对于等长码来说，如果原等长码是非奇异的，则它的任意 L 次扩展码也要求是非奇异的，这是唯一可译性所要求的。

现在考虑对单符号信源 S 的 L 次扩展信源 $S^{(L)}$进行等长编码，扩展信源 $S^{(L)}$含 K^L 个消息，要得到长为 n 的唯一可译码，必须满足

$$K^L\leqslant D^n \tag{3-5}$$

式中，K 为信源符号集$\{a_1, a_2, \cdots, a_K\}$的元素个数，$L$ 为信源输出序列长度，D 为码符号集$\{b_1, b_2, \cdots, b_D\}$的元素个数，$n$ 为码字长度。

对式（3-5）两边取对数并整理，得

$$\frac{n}{L}\geq\frac{\log K}{\log D} \tag{3-6}$$

式中，$\frac{n}{L}$表示信源序列中平均每个信源符号所需要的码符号数，式（3-6）说明对于等长唯一可译码，平均每个信源符号至少需要$\frac{\log K}{\log D}$个码符号数来对其进行编码变换。

对例 3.7 中所列举的英文 27 个字符进行等长二进制编码，即 K=27，L=1，D=2，则根据式（3-6），要得到唯一可译码，应满足码长 $n \geq \frac{\log K}{\log D} \times L = \frac{\log 27}{\log 2} \times 1 = 4.75$，实际码长 n 应取整数 5，即每个英文符号都需要 5 位二进制数进行编码，这 5 位二进制数在等概情况下最多可载有 5 比特信息量。实际上，由于英文字符各字符出现的概率不同，以及各字符之间的关联性（如 in 后面出现 g 的概率极大），已知的东西多，得到的信息量就小，因此收到一个 5 位二进制数后得到的信息量将大大小于 5 比特，这样的传输效率是很低的。

可以设想，对于那些出现概率极小的字符序列不予编码，可以减小平均码长 $\bar{n}$，从而提高信息传输率 R_D（$=H(X)/\bar{n}$），当然这样会引起一定的失真。下面的定理 3.1 将证明，当满足一定的条件时，在 $L \to \infty$ 时，失真 $p_e \to 0$。

定理 3.1 等长编码定理 设离散无记忆信源 $S=\{x_1, x_2, \cdots, x_k\}$ 的熵为 $H(X)$，S 的 L 维扩展信源为 $S^{(L)}=\{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_{k^L}\}$，对信源输出的 L 长序列 $\boldsymbol{s}_i$，$i=1, 2, \cdots, K^L$ 进行等长编码，码字是长度为 n 的 D 进制符号串，若满足条件 $\frac{n}{L} > \frac{H(X)+\varepsilon}{\log D}$，则 $L \to \infty$ 时，可使译码差错 $p_e < \delta$（ε, δ 为无穷小量）；反之，若 $\frac{n}{L} < \frac{H(X)+\varepsilon}{\log D}$，则不可能实现无差错编码。

该定理的证明冗长烦琐，此处不予证明。

现在将定理要求的下界 $\frac{n}{L} > \frac{H(X)+\varepsilon}{\log D}$ 与式（3-6）要求的下界 $\frac{n}{L} \geq \frac{\log k}{\log D}$ 进行比较，$H(X)$ 是单符号信源 $S=\{x_1, x_2, \cdots, x_k\}$ 的熵，根据极大离散熵定理，信源等概分布时熵值最大，其最大值为 $\log k$，即有 $H(X) \leqslant \log K$，这就说明了定理 3.1 要求的下界比式（3-6）要求的下界更紧致，对 $\frac{n}{L}$ 的要求更低。

仍然考虑例 3.7 中所列举的英文字符信源，根据式（3-6），可求出

$$n_1 \geq L\frac{\log K}{\log D} = 1 \times \frac{\log 27}{\log 2} = 4.75$$

根据定理 3.1 的要求，可求出

$$n_2 \geq L\frac{H(X)}{\log D} = 1 \times \frac{4.03}{\log 2} = 4.03$$

显然，$n_1 > n_2$，实际上，由于要求码长取整数，故只能取 $n_1 = n_2 = 5$。

定理 3.1 要求 $\frac{n}{L} > \frac{H(X)+\varepsilon}{\log D}$，即 $1 > \frac{L[H(X)+\varepsilon]}{n \log D}$，可看出比值 $\frac{LH(X)}{n \log D}$ 是一个小于 1 的无量纲纯数，定义它为等长编码的编码效率，记为

$$\eta = \frac{LH(X)}{n \log D} \tag{3-7}$$

编码效率 η 是衡量编码质量的一个重要指标，对信源编码时应尽量提高编码效率。

【例 3.9】 （1）给定离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & x_2 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}$，对该信源进行二进制等长编码，先确定码长，在此例中，信源消息数 $K=3$，信源序列长 $L=1$，码符号数 $D=2$，根据式（3-6）可算得

$$n_1 \geq L\frac{\log K}{\log D} = 1\times\frac{\log 3}{\log 2} = 1.585 \tag{3-8}$$

根据定理 3.1，要求码长 $n_2 \geq L\dfrac{H(X)}{\log D}$，先计算出信源熵

$$H(X) = -0.1\log 0.1 - 0.7\log 0.7 - 0.2\log 0.2 = 1.156\text{（比特/符号）}$$

$$n_2 \geq L\frac{H(X)}{\log D} = 1\times\frac{1.116}{\log 2} = 1.116 \tag{3-9}$$

显然，定理 3.1 要求的码长下界比式（3-6）要求的码长下界更紧致，但由于码长必须是整数，根据式（3-8）和式（3-9）都应取 $n_1 = n_2 = 2$。

可将信源编码为 a_0: 00，a_1: 01，a_2: 10，得到的是唯一可译码，编码差错率为零。根据式（3-7）计算编码效率

$$\eta_1 = \eta_2 = \frac{LH(X)}{n\log D} = \frac{1\times 1.155}{2\log 2} = 0.587$$

（2）现在考虑对原信源进行 L 维扩展，为计算简便，取 $L=2$，得到新信源

$$\begin{bmatrix} X^{(2)} \\ q(X^{(2)}) \end{bmatrix} = \begin{bmatrix} x_0x_0 & x_0x_1 & x_0x_2 & x_1x_0 & x_1x_1 & x_1x_2 & x_2x_0 & x_2x_1 & x_2x_2 \\ 0.01 & 0.07 & 0.02 & 0.07 & 0.49 & 0.14 & 0.02 & 0.14 & 0.04 \end{bmatrix}$$

对扩展信源进行二进制等长编码

① 先根据式（3-6）确定码长 n_3，对扩展信源有 $K=3$，$L=2$，$D=2$，可算得

$$n_3 \geq L\frac{\log K}{\log D} = 2\times\frac{\log 3}{\log 2} = 3.170$$

由于码长必须是整数，取 $n_3=4$，对扩展信源的编码见表 3-6。

表 3-6 取 $n_3=4$ 对扩展信源的编码

信源序列	u_0u_0	u_0u_1	u_0u_2	u_1u_0	u_1u_1	u_1u_2	u_2u_0	u_2u_1	u_2u_2
码　字	0000	0001	0010	0011	0100	0101	0110	0111	1000

信源序列与码字是一一对应的，得到的是唯一可译码，编码差错率为零。

算得编码效率

$$\eta_3 = \frac{LH(X)}{n_3\log D} = \frac{2\times 1.156}{4\log 2} = 0.578$$

② 再根据定理 3.1 确定码长 n_4，有

$$n_4 \geq L\frac{H(X)}{\log D} = 2\times\frac{1.156}{\log 2} = 2.312$$

取整数 $n_4=3$，对扩展信源的编码见表 3-7。

表 3-7　取 $n_4=3$ 对扩展信源的编码

信源序列	u_0u_0	u_0u_1	u_0u_2	u_1u_0	u_1u_1	u_1u_2	u_2u_0	u_2u_1	u_2u_2
概　率	0.01	0.07	0.02	0.07	0.49	0.14	0.02	0.14	0.04
码　字	000	001	000	010	011	100	101	110	111

对于二进制码，取码长为 3，共可构成 $2^3=8$ 个不同的码字，而扩展信源含 9 个序列，所以编码时对信源输出序列中出现概率最小的 u_0u_0 和 u_0u_2 赋予同一个码字 000，这样势必带来编码误差 p_e，可算出

$$p_e=(0.01+0.02)\times\left[\frac{0.01}{0.01+0.02}+\frac{0.02}{0.01+0.02}\right]=0.03$$

算出这种情况下的编码效率

$$\eta_4=\frac{LH(X)}{n_4\log D}=\frac{2\times1.156}{3\log 2}=0.771$$

较之 $n_2=2$ 的情况，编码效率是原值的 $\dfrac{\eta_4}{\eta_2}=\dfrac{0.771}{0.587}=1.31$（倍）。

显然信息传输率提高了许多，当然上面两种情况下的编码效率都是不高的。随着 L 继续增大，编码效率将进一步提高，但实际上，对于等长码，往往要在 L 很大的情况下，才能满足对编码效率及编码差错率的要求。

定理 3.1 要求 $\dfrac{n}{L}>\dfrac{H(X)}{\log D}$，这是为了实现无差错编码每个信源符号所需要的最少码符号数，这是等长码编码时的理论极限。事实上这种几乎无失真的编码代价是，要求信源序列长 $L\to\infty$，L 大就意味着实现的复杂性及译码的时延加大。

这条定理是在离散无记忆的条件下证得的，它也同样适用于平稳有记忆信源，但要求有记忆信源的极限熵 $H_\infty(X)$及极限方差 $\sigma_\infty^2(X)$ 存在，定理中的 $H(X)$改为 $H_\infty(X)$即可。

3.3　变长码及变长编码定理

3.3.1　变长码

3.2 节对于等长码的讨论是在 L 足够大的条件下得到的结论，L 太大，给编码、译码带来很大的困难，且设备复杂性增加；而当 L 为有限值时，又总会带来一定程度的失真。对于变长码，往往在 L 不太大的情况下就可编出高效且无失真的码。

变长码也必须是唯一可译码，要满足唯一可译性，则要求原码的任意 L 次扩展码也是唯一可译的。由图 3-5 的码分类结构图可看出，在唯一可译码中，等长码在 3.2 节已经讨论过了，本节讨论变长码，变长码又分为即时码和延长码，为保证即时译码，要求变长唯一可译码采用即时码。

对于变长码，不要求每个码字的码长一致，但要使整个码集的平均码长力求最小，此时编码效率最高，信源的冗余度得到最大程度的压缩，从而提高了信息传输率，而这正是信源编码的主要关注点。对于给定信源，使平均码长达到最小的编码方法，称为最佳编码，得到的码集称为最佳码。

信源无失真变长编码定理给出了在无失真编码前提下，平均码长的上界及下界。

在讲述定理以前，先给出克拉夫特不等式。

3.3.2 克拉夫特不等式

如上所述，唯一可译码可保证无差错译码，下面的定理 3.2 对唯一可译码的码长做出了要求。

定理 3.2 D 进制码字集合 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$，码集中每一 $\boldsymbol{c}_m$（$m=1, 2, \cdots, M$）都是一个 D 进制符号串，设 $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M$ 对应的码长分别是 $n_1, n_2, \cdots, n_M$，则存在唯一可译码的充要条件是

$$\sum_{m=1}^{M} D^{-n_m} \leq 1 \tag{3-10}$$

式（3-10）也称克拉夫特不等式，该不等式在 1949 年由 Kraft（克拉夫特）提出，它是在即时码的条件下证明的，1956 年 McMillan（麦克米伦）证得该不等式对唯一可译码也成立，1961 年 Karush（卡拉什）简化了它的证明。

证明：

① 必要性。设码 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$唯一可译，$n_m$ 是码字 $\boldsymbol{c}_m$（$m=1, 2, \cdots, M$）所对应的码长，对原码 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$进行 L 次扩展（L 为任意正整数），考虑下面的求和式

$$\left(\sum_{m=1}^{M} D^{-n_m}\right)^{L} = \left(\sum_{m_1=1}^{M} D^{-n_{m_1}}\right)\cdot\left(\sum_{m_2=1}^{M} D^{-n_{m_2}}\right)\cdots\left(\sum_{m_L}^{M} D^{-n_{m_L}}\right)$$

$$= \sum_{m_1=1}^{M}\sum_{m_2=1}^{M}\cdots\sum_{m_L=1}^{M} D^{-(n_{m_1}+n_{m_2}+\cdots+n_{m_L})} \tag{3-11}$$

等式（3-11）右边包括 M^L 项，每一项都对应着原码 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$的 L 次扩展码 $\boldsymbol{C}^{(L)}$ 的一个码序列，

M^L 正是 L 次扩展码 $\boldsymbol{C}^{(L)}$ 的码序列总数。

记 L 次扩展码的一个码序列 $\boldsymbol{c}_i^{(L)}$（$i=1, 2, \cdots, M^L$）由原码中的 L 个码字 $\boldsymbol{c}_{i1}, \boldsymbol{c}_{i2}, \cdots, \boldsymbol{c}_{iL}$ 并列而成，$n_{i1}, n_{i2}, \cdots, n_{iL}$ 分别是各个码字 $\boldsymbol{c}_{i1}, \boldsymbol{c}_{i2}, \cdots, \boldsymbol{c}_{iL}$ 对应的码长，则 $\boldsymbol{c}_i^{(L)}$ 的码元长度 l 为

$$l = n_{i1} + n_{i2} + \cdots + n_{iL}$$

因为原码的长度 $n_{i1}, n_{i2}, \cdots, n_{iL}$ 从$\{n_1, n_2, \cdots, n_M\}$中取值，故 l 也是一个变量，记原码中，码长的最大值和最小值分别为 $n_{\max}$ 和 $n_{\min}$，即

$$\begin{cases} n_{\max} = \max\{n_1, n_2, \cdots, n_M\} \\ n_{\min} = \min\{n_1, n_2, \cdots, n_M\} \end{cases}$$

则 l 的取值范围满足 $Ln_{\min} \leqslant l \leqslant Ln_{\max}$，一般取 $n_{\min}=1$，则有

$$L \leqslant l \leqslant Ln_{\max} \tag{3-12}$$

设 L 次扩展码的所有码序列中，码长为 l 的序列有Δ_l个，则根据式（3-11）及式（3-12），下式成立

$$\left(\sum_{m=1}^{M} D^{-n_m}\right)^{L} = \sum_{l=L}^{Ln_{\max}} \Delta_l D^{-l} \tag{3-13}$$

因为原码是唯一可译码，由码的唯一可译性，知它的 L 次扩展码也是唯一可译码，又因为长度为 l 的 D 进制序列有 D^l 个，要满足唯一可译性，则必有 $\Delta_l \leqslant D^l$，即

$$\Delta_l D^{-l} \leqslant 1 \tag{3-14}$$

将式（3-14）代入式（3-13），得

$$\left(\sum_{m=1}^{M} D^{-n_m}\right)^L \leq \sum_{l=L}^{Ln_{\max}} 1 = Ln_{\max} - (L-1) \leq Ln_{\max}$$

$$\sum_{m=1}^{M} D^{-n_m} \leq (Ln_{\max})^{\frac{1}{L}} \tag{3-15}$$

当 $L \to \infty$ 时，式（3-15）右边 $\lim_{L\to\infty}(Ln_{\max})^{\frac{1}{L}} = 1$，将这一结果代回式（3-15），就得到

$$\sum_{m=1}^{M} D^{-n_m} \leq 1$$

② 充分性。设不等式 $\sum_{m=1}^{M} D^{-n_m} \leq 1$ 成立，因为定理成立是在满足式（3-10）的条件下，总可找到至少一种唯一可译码。下面采用构造性证明，即在满足式（3-10）的条件下，构造出一种唯一可译码。

记整数 $h_i \geqslant 1$（$i = 0, 1, \cdots, j-1$），不失一般性，假设 $h_0 < h_1 < \cdots < h_{j-1}$，定义

$$\begin{cases} w_0 = 0 \\ w_j = \sum_{i=0}^{j-1} D^{-h_i} \quad (j = 1, 2, \cdots, M) \end{cases}$$

根据假设 $\sum_{m=1}^{M} D^{-n_m} \leq 1$，$w_j$ 是一个 h_j 位的 D 进制小数。

构造码：取 $w_j = \sum_{i=0}^{j-1} D^{-h_i}$ 的 h_j 位小数，构造符号串 $\boldsymbol{c}_j$（j=1, 2, …, M），这样得到的码 $\boldsymbol{C}$={$\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_M$}是唯一可译的。

下面证明码 $\boldsymbol{C}$={$\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_M$}确实是唯一可译的。

用反证法，设该码不唯一可译，其中有一个码字 $\boldsymbol{c}_s$ 是另一个码字 $\boldsymbol{c}_k$ 的字头，若 w_s 是 $\boldsymbol{c}_s$ 对应的 h_s 位小数，w_k 是 $\boldsymbol{c}_k$ 对应的 h_k 位小数，因为 $\boldsymbol{c}_s$ 是 $\boldsymbol{c}_k$ 的字头，则小数 w_k 大于小数 w_s，记为

$$w_k > w_s \tag{3-16}$$

将式（3-16）两边同时乘以 D^{h_s}，得 $w_k D^{h_s} > w_s D^{h_s}$。

乘以 D^{h_s} 后，就将小数 w_k 和 w_s 的小数点的位置各向右移了 h_s 位，使 $w_s D^{h_s}$ 恰好为一整数，而 $w_k D^{h_s}$ 除整数部分外，还有小数，如图 3-6 所示。

显然有

$$w_k D^{h_s} < w_s D^{h_s} + 1$$

因此得

$$(w_k - w_s) D^{h_s} < 1 \tag{3-17}$$

另一方面，根据码的构造法，有

$$(w_k - w_s) D^{h_s} = \left(\sum_{i=0}^{k-1} D^{-h_i} - \sum_{i=0}^{s-1} D^{-h_i}\right) D^{h_s}$$

$$=\left(\sum_{i=s}^{k-1}D^{-h_i}\right)D^{h_s}$$
$$>(D^{-h_s})D^{h_s}=1$$

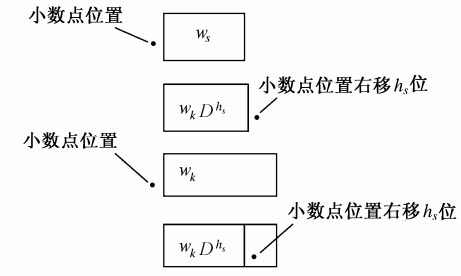


图 3-6　小数点右移示意图

即

$$(w_k - w_s)D^{h_s} > 1 \qquad (3\text{-}18)$$

式（3-17）和式（3-18）矛盾，说明原假设是错误的，构造的码 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$必唯一可译。

证毕

定理 3.2 只是说明$\sum_{m=1}^{M}D^{-n_m}\leq 1$是存在唯一可译码的充要条件，这里强调的是“存在”，但它并不是唯一可译码的充要条件，换言之，唯一可译码一定满足克拉夫特不等式，反之，满足克拉夫特不等式的码不一定是唯一可译码。

下面验证表 3-2 中所列的几种码是否满足克拉夫特不等式。

码 1 不是唯一可译码，可计算出　$\sum_{m=1}^{M}D^{-n_m}=2^{-1}+2^{-1}+2^{-2}+2^{-2}=\frac{3}{2}>1$

码 3 是唯一可译码，可计算出　$\sum_{m=1}^{M}D^{-n_m}=2^{-1}+2^{-2}+2^{-3}+2^{-4}=\frac{15}{16}<1$

有些非唯一可译码也满足克拉夫特不等式，例如，码 $\boldsymbol{C}$= {1，11，111，1111}就不是唯一可译码，但存在$\sum_{m=1}^{M}D^{-n_m}=2^{-1}+2^{-2}+2^{-3}+2^{-4}=\frac{15}{16}<1$。

3.3.3　变长编码定理

信源熵 $H(X)$从它的物理意义来说，是每个信源符号平均所携带的信息量，设信源含有 M 个消息 $x_1, x_2, \cdots, x_M$，其概率统计模型为$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$，通过信源编码后，设每个消息对应的码长为 $n_1, n_2, \cdots, n_M$，每个信源符号平均需要$\overline{n}=\sum_{m=1}^{M}q(x_m)n_m$ 个码符号对其编码，$H(X)/\overline{n}$ 是平均每个码符号所携带的信息量。码字在信道传输过程中，若时间以码元时间（传输一个码符号的时间）为单位，$H(X)/\overline{n}$ 就是前面所定义的信息传输率 R_D。由此看出，平均码长$\overline{n}$越短，则信息传输率越高，所以从提高传输效率来考虑，对信源编码时，尽可

能选平均码长小的码。给定了信源及码符号集后，在所有的唯一可译码中，平均码长最小的码称为紧致码（最佳码），信源编码问题就是寻找最佳码。

如上所述，在保证无失真的前提下，信源编码希望平均码长尽可能小，下面的定理给出了唯一可译码的码长取值范围。

定理 3.3 给定熵为 $H(X)$的离散无记忆信源，及有 D 个元素的码符号集，则总可找到一种无失真编码方法，构成唯一可译码，其平均码长 $\bar{n}$ 满足

$$\frac{H(X)}{\log D} \leq \bar{n} < 1 + \frac{H(X)}{\log D} \tag{3-19}$$

证明：

（1）先证下界 $\bar{n} \geq \dfrac{H(X)}{\log D}$，即证 $H(X) - \bar{n}\log D \leq 0$。

设离散无记忆信源含 M 个消息，信源熵为 $H(X)$，其统计模型为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$$

则

$$\begin{aligned} & H(X) - \bar{n}\log D \\ &= \sum_{m=1}^{M} q(x_m)\log\frac{1}{q(x_m)} - \log D \cdot \sum_{m=1}^{M} q(x_m) n_m \\ &= \sum_{m=1}^{M} q(x_m)\log\frac{1}{q(x_m)} + \sum_{m=1}^{M} q(x_m)\log D^{-n_m} \\ &= \sum_{m=1}^{M} q(x_m)\log\frac{D^{-n_m}}{q(x_m)} \\ &\leq \log\sum_{m=1}^{M} q(x_m)\frac{D^{-n_m}}{q(x_m)} \\ &= \log\sum_{m=1}^{M} D^{-n_m} \\ &\leq \log 1 = 0 \qquad \text{（利用克拉夫特不等式）} \end{aligned}$$

当 $\dfrac{D^{-n_m}}{q(x_m)} = 1$，即 $q(x_m) = D^{-n_m} = \left(\dfrac{1}{D}\right)^{n_m}$ 时，上式等号成立，有

$$H(X) - \bar{n}\log D = 0$$

$$\bar{n} = \frac{H(X)}{\log D} \quad \text{（下界）}$$

这时得到的码就是紧致码，意味着信源消息概率分布 $q(x_m)$必须有$\left(\dfrac{1}{D}\right)^{h_m}$（$h_m$ 为整数）的形式，直接取各消息码字的码长 n_m 等于 $q(x_m)$所对应的指数 h_m 即可。

这就是例 3.8 所列举的情况，在例 3.8 中，信源分布可以表示为

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ (1/2)^2 & (1/2)^2 & (1/2)^3 & (1/2)^3 & (1/2)^4 & (1/2)^4 & (1/2)^4 & (1/2)^4 \end{bmatrix}$$

取信源各消息相应的码字的码长等于其分布概率所对应的指数，即

$$n_0=2，\ n_1=2，\ n_2=3，\ n_3=3，\ n_4=4，\ n_5=4，\ n_6=4，\ n_7=4$$

得到的信源编码就是紧致码（最佳码）。

（2）再证上界：$\overline{n}<1+\dfrac{H(X)}{\log D}$

采用构造性证明，只要构造一种唯一可译码，满足 $\overline{n}<1+\dfrac{H(X)}{\log D}$ 即可。

先将概率 $q(x_m)$写成 $q(x_m)=\left(\dfrac{1}{D}\right)^{t_m}$ 的形式（t_m 不一定为整数）。

构造码，使码长满足

$$\begin{cases} n_m=t_m & t_m \text{为整数时} \\ t_m<n_m<t_m+1 & t_m \text{不是整数，但在（}t_m\text{，}t_m\text{+1）之间总有一整数} \end{cases}$$

综合两种情况，码长 n_m 满足

$$t_m \leqslant n_m < t_m+1 \tag{3-20}$$

由 $q(x_m)=\left(\dfrac{1}{D}\right)^{t_m}$，可解出

$$t_m=\frac{-\log q(x_m)}{\log D} \tag{3-21}$$

将式（3-21）代入式（3-20）右边的不等式 $n_m<t_m+1$，即有

$$n_m<-\frac{\log q(x_m)}{\log D}+1 \tag{3-22}$$

对式（3-22）两边求统计平均值得

$$\sum_{m=1}^{M}q(x_m)\cdot n_m<-\sum_{m=1}^{M}q(x_m)\left[\frac{\log q(x_m)}{\log D}-1\right]$$

即 $$\overline{n}<1+\frac{H(X)}{\log D}$$ 证毕

定理 3.3 说明，只有满足 $\overline{n}\geq\dfrac{H(X)}{\log D}$，才能构成唯一可译码，否则唯一可译码不存在。但平均码长 $\overline{n}$ 应小于 $\left[1+\dfrac{H(X)}{\log D}\right]$，这是应 $\overline{n}$ 应尽可能短的要求，这时得到的码是最佳码，其实 $\overline{n}>1+\dfrac{H(X)}{\log D}$ 也能找到唯一可译码。

【例 3.10】 信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/2 & (1/2)^2 & (1/2)^3 & (1/2)^3 \end{bmatrix}$，对信源进行二进制变长编码，$D=2$，信源各消息概率恰好表示成 $D=2$ 的整数次幂，取码长等于其幂次，即取 n_1=1，n_2=2，n_3=3，n_4=3 对信源各消息编码，得到的码就是紧致码，该码送入二进制信道传输，下面计算信息传输率 R_D。

$$\overline{n}=\sum_m n_m q(x_m)=1\times\frac{1}{2}+2\times\frac{1}{4}+2\times3\times\frac{1}{8}=1.75\text{（码元/符号）}$$

信源熵 $H(X)=-\sum_{i=1}^{4}q(x_i)\log q(x_i)=\frac{1}{2}\log 2+\frac{1}{4}\log 4+2\times\frac{1}{8}\log 8=1.75\log 2=1.75$（比特/符号）

$$R_D=\frac{H(X)}{\bar{n}}=1$$（比特/码元时间）

因为信息传输率 R_D 的值小于等于 1，所以上述 R_D =1 达到最大值，这种编码法得到的码集为紧致码。

【例 3.11】 对下述信源进行二进制变长编码，即

$$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.2 & 0.1 & 0.3 & 0.15 & 0.25 \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ (1/2)^{2.322} & (1/2)^{3.322} & (1/2)^{1.737} & (1/2)^{2.733} & (1/2)^{2} \end{bmatrix}$$

根据式（3-20），即码长 n_m 应满足 $t_m \leqslant n_m < t_m+1$，$t_m$ 是消息 x_m（m =1，2，3，4，5）的 2 次幂概率所对应的幂次，取 $\{x_1, x_2, x_3, x_4, x_5\}$ 所对应的码字的码长分别为 n_1=3，n_2=4，n_3=2，n_4=3，n_5=2，计算出平均码长

$$\bar{n}=0.2\times 3+0.1\times 4+0.3\times 2+0.15\times 3+0.25\times 2=2.55$$

信源熵 $H(X)=-0.2\log 0.2-0.1\log 0.1-0.3\log 0.3-0.15\log 0.15-0.25\log 0.25$

$=2.228$（比特/符号）

则有 $$\frac{H(X)}{\log D}=\frac{H(X)}{\log 2}=2.228$$

满足式（3-19） $$\frac{H(X)}{\log D}\leqslant \bar{n}< 1+\frac{H(X)}{\log D}$$

将定理 3.3 推广到 L 次扩展信源，就得到下述定理。

定理 3.4 变长编码定理（Shannon 第一定理）

给定熵为 $H(X)$的离散无记忆信源$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$，其 L 次扩展信源 $\begin{bmatrix} \boldsymbol{X} \\ q(\boldsymbol{X}) \end{bmatrix}=\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_M \\ q(\boldsymbol{x}_1) & q(\boldsymbol{x}_2) & \cdots & q(\boldsymbol{x}_{M^L}) \end{bmatrix}$的熵记为 $H(\boldsymbol{X})$，给定有 D 个元素的码符号集，对扩展信源进行编码，总可以找到一种唯一可译码，使码长 $\bar{n}_L$ 满足

$$\frac{H(X)}{\log D}\leqslant \frac{\bar{n}_L}{L}< \frac{H(X)}{\log D}+\frac{1}{L} \qquad (3\text{-}23)$$

式中，L 为信源序列长度。

记扩展信源输出序列为 $\boldsymbol{x}_i$ (i =1, 2, …, M^L)，出现概率为 $q(\boldsymbol{x}_i)$，对应的码字为 $\boldsymbol{c}_i$，其码长为 n_i，则 $\bar{n}_L=\sum_{i=1}^{M^L}q(\boldsymbol{x}_i)n_i$ 为码集 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_{M^L}\}$ 的平均码长。

记 $\frac{\bar{n}_L}{L}=\bar{n}$ 为信源每个符号所对应的平均码字数，则式（3-23）为

$$\frac{H(X)}{\log D}\leqslant \bar{n}< \frac{H(X)}{\log D}+\frac{1}{L} \qquad (3\text{-}24)$$

当 L→∞时，有 $\bar{n}\to\frac{H(X)}{\log D}$。

显然，定理 3.3 是定理 3.4 在 L =1 时的特例。

证明：将定理 3.3 用于扩展信源$\begin{bmatrix} \boldsymbol{X} \\ q(\boldsymbol{X}) \end{bmatrix}=\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_M \\ q(\boldsymbol{x}_1) & q(\boldsymbol{x}_2) & \cdots & q(\boldsymbol{x}_{M^L}) \end{bmatrix}$，得

$$\frac{H(\boldsymbol{X})}{\log D} \leqslant \bar{n}_L < \frac{H(\boldsymbol{X})}{\log D}+1 \tag{3-25}$$

根据式（2-43），对于离散无记忆信源，有 $H(\boldsymbol{X})= L\,H(X)$，代入式（3-25），得

$$\frac{H(X)}{\log D} \leqslant \frac{\bar{n}_L}{L} < \frac{H(X)}{\log D}+\frac{1}{L}$$

$$\frac{H(X)}{\log D} \leqslant \bar{n} < \frac{H(X)}{\log D}+\frac{1}{L}$$

显然，当 $L\to\infty$ 时，式（3-25）变为 $\frac{H(X)}{\log D} \leqslant \bar{n} < \frac{H(X)}{\log D}$，即

$$\lim_{L\to\infty} \bar{n} = \frac{H(X)}{\log D}$$

可见，在 L→∞的极限情况下，信息传输率 $R_D = \frac{H(U)}{\bar{n}} = \log D$，这就是 R_D 所能达到的极限值，它对应于等概分布。 证毕

因此，Shannon 第一定理的物理意义在于：对信源进行编码，使编码后的码集中各码字尽可能等概分布，如果将这码集看成一个新的信源，则这时新信源所含信息量最大。

$\bar{n}$ 总是大于或等于 $\frac{H(X)}{\log D}$，为了衡量各种编码是否已达到理想情况，用二者的比值来定义**编码效率**，即

$$\eta \triangleq \frac{H(X)/\log D}{\bar{n}} = \frac{H(X)}{\bar{n}\log D} \tag{3-26}$$

式中，η 是一个无量纲的数，一般情况下 η<1，在极限情况下 η=1。

将式（3-26）定义的变长编码的编码效率与式（3-7）定义的等长编码的编码效率 $\eta = \frac{LH(X)}{n\log D} = \frac{H(X)}{(n/L)\log D}$ 进行比较，可看出，式（3-7）中的 n/L 就是信源每个符号所对应的平均码字数，所以等长编码和变长编码这两种情况下定义的编码效率是一致的。

【例 3.12】

（1）计算例 3.10 中所得编码的编码效率 η。

例 3.10 已算出：信源熵 $H(X)$ =1.75（比特/信源符号），平均码长 $\bar{n}$ =1.75（码符号/信源符号），又有 logD = log2 = 1（比特/码符号），则可算出编码效率 $\eta = \frac{H(X)}{\bar{n}\log D} = \frac{1.75}{1.75\times\log 2} = 1$，这是最佳情况。

（2）计算例 3.11 中所得编码的编码效率 η。

例 3.11 已算出：信源熵 $H(X)$=2.228（比特/信源符号），平均码长 $\bar{n}$ = 2.55（码符号/信源符号），又有 logD = log2 = 1（比特/码符号），则可算出编码效率 $\eta = \frac{H(X)}{\bar{n}\log D} = \frac{2.228}{2.55\times\log 2} = 0.87$，显然较理想情况为低。

【例 3.13】　给出一维信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ 3/4 & 1/4 \end{bmatrix}$，可算出信源熵

$$H(X)=\frac{1}{4}\log 4+\frac{3}{4}\log\frac{4}{3}=0.811$$

取 $D=2$ 对信源进行二进制编码，取 x_1: 0, x_2: 1，则编码效率 $\eta_1=\dfrac{H(X)}{\bar{n}\log D}=\dfrac{0.811}{1\times\log 2}=0.811$。

根据 Shannon 第一定理，L 增大时，η也随之增大，下面对 $L=2$ 的扩展信源 $\begin{bmatrix} \boldsymbol{X} \\ q(\boldsymbol{X}) \end{bmatrix}=\begin{bmatrix} x_1x_1 & x_1x_2 & x_2x_1 & x_2x_2 \\ 9/16 & 3/16 & 3/16 & 1/16 \end{bmatrix}=\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \boldsymbol{x}_4 \\ 9/16 & 3/16 & 3/16 & 1/16 \end{bmatrix}$进行变长编码。

遵照概率大码长小的原则，使平均码长 $\bar{n}$ 尽可能小，取 x_1x_1: 0, x_1x_2: 10, x_2x_1: 110, x_2x_2: 111 算出平均码长

$$\bar{n}=\frac{1}{L}\sum_{i=1}^{4}n(\boldsymbol{x}_i)q(\boldsymbol{x}_i)=\frac{1}{2}\times\left(1\times\frac{9}{16}+2\times\frac{3}{16}+3\times\frac{3}{16}+3\times\frac{1}{16}\right)=\frac{27}{32}$$

则编码效率

$$\eta_2=\frac{H(X)}{\bar{n}\log D}=\frac{0.811}{(27/32)\log 2}=0.961$$

显然 $\eta_2>\eta_1$，编码效率提高了许多，可见对于变长码，L 不需要很大就可以达到相当高的编码效率。

3.4　变长码的编码方法

常用的变长编码法有如下三种：香农（Shannon）编码法、费诺（Fano）编码法、霍夫曼（Huffman）编码法。

对于同一种信源，三种编码法中以香农编码法的编码效率最低，但这种编码法对于证明变长编码定理起了很重要的作用，所以它有着重要的理论指导意义。

费诺编码法也不是一种最佳编码法，但用这种方法有时候也能找到紧致码。

一般情况下，对于同一信源，三种编码法以霍夫曼编码法得到的平均码长 $\bar{n}$ 最短，即编码效率 $\eta=\dfrac{H(X)}{\bar{n}\log D}$ 最高。

下面逐一介绍这三种编码法。

3.4.1　香农编码法

实际上在证明定理 3.3 时就介绍过一种唯一可译码的编码方法如下。

记离散信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$，给定有 D 个元素的码符号集，对信源进行变长编码，将各消息概率 $q(x_m)$（$m=1, 2, \cdots, M$）写成如下形式，有

$$q(x_m)=\left(\frac{1}{D}\right)^{t_m} \tag{3-27}$$

式中，t_m 不一定为整数，取码长 n_m（$m=1, 2, \cdots, M$）满足

$$t_m \leqslant n_m < t_m + 1 \tag{3-28}$$

由式（3-27）得 $t_m = -\dfrac{\log q(x_m)}{\log D}$，代入式（3-28）即有

$$-\frac{\log q(x_m)}{\log D} \leqslant n_m < -\frac{\log q(x_m)}{\log D} + 1 \tag{3-29}$$

对于二进制编码，$D=2$，由式（3-29）得

$$-\log q(x_m) \leqslant n_m < -\log q(x_m) + 1 \tag{3-30}$$

式（3-30）就是二进制香农编码法其码长的取值范围。

香农编码法具体步骤如下，参见例 3.14 中的表 3-8。

① 将信源发出的 M 个消息，按其概率递减顺序进行排列，得

$$q(x_1) \geqslant q(x_2) \geqslant q(x_3) \geqslant \cdots \geqslant q(x_M)$$

② 计算出各消息的$-\log q(x_m)$值，$m=1, 2, \cdots, M$；

③ 根据式（3-30）：$-\log q(x_m) \leqslant n_m < -\log q(x_m) + 1$（$-\log q(x_m)$为整数时取等号），计算出每个消息的二进制代码的长度 n_m（$m=1, 2, \cdots, M$），n_m 取正整数；

④ 为得到唯一可译码，计算出第 m 个消息的累加概率 $p_m = \sum_{i=1}^{m-1} q(x_i)$，再将 p_m 变换成二进制小数，取小数点后面 n_m 位作为第 m 个消息的代码组（码字）。

【例 3.14】 对给定信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$ 进行 $D=2$ 的二进制香农编码。

下面以消息 x_5 为例，对其进行编码。

计算出$-\log q(x_5) = -\log 0.15 = 2.74$，取整数 $n_5 = 3$ 作为 x_5 的码字的码长，计算出消息 x_1, x_2, x_3, x_4 的累加概率，有

$$p_5 = \sum_{i=1}^{5-1} q(x_i) = 0.2 + 0.19 + 0.18 + 0.17 = 0.74$$

将 0.74 变换成二进制小数 $(0.74)_{10} = (0.1011110)_2$，取小数点后面 3 位 101 作为 x_5 的代码。其余消息的代码也可以用同样的方法计算得到，计算结果列于表 3-8 中。

表 3-8　香农编码

消息符号 x_m	消息概率 $q(x_m)$	$-\log q(x_m)$	码长 n_m	累加概率 $p_m = \sum_{i=1}^{m-1} q(x_i)$	码字 $\boldsymbol{c}_m$
x_1	0.2	2.34	3	0	000
x_2	0.19	2.41	3	0.2	001
x_3	0.18	2.48	3	0.39	011
x_4	0.17	2.56	3	0.57	100
x_5	0.15	2.74	3	0.74	101
x_6	0.10	3.34	4	0.89	1110
x_7	0.01	6.66	7	0.99	1111110

下面计算该编码的编码效率 $\eta = \dfrac{H(X)}{\overline{n} \log D}$。

先算出信源熵

$$H(X)=-\sum_{m=1}^{7}q(x_m)\log q(x_m)$$

$$=-0.2\log0.2-0.19\log0.19-0.18\log0.18-0.17\log0.17-0.15\log0.15-0.10\log0.10-0.01\log0.01$$

$$=2.61\text{（比特/符号）}$$

平均码长 $\bar{n}=\sum_{m=1}^{7}n_m q(x_m)$

$$=3\times(0.2+0.19+0.18+0.17+0.15)+4\times0.10+7\times0.01$$

$$=3.14\text{（比特/符号）}$$

则编码效率 $\eta=\dfrac{H(X)}{\bar{n}\log D}=\dfrac{2.61}{3.14\times1}=0.831$

3.4.2 费诺编码法

香农第一定理的物理意义阐明：信源等概分布时，信息传输率最大，应用到对信源进行编码，应使编码后的码集尽可能等概分布，如果将该码集看成一个新的信源，这时新信源所含信息量最大。

记离散信源 $\begin{bmatrix}X\\q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_m & \cdots & x_M\\q(x_1) & q(x_2) & \cdots & q(x_m) & \cdots & q(x_M)\end{bmatrix}$，给定有 D 个元素的码符号集，对信源进行编码，根据香农第一定理的指导思想，费诺编码法的具体步骤如下，参见例 3.15 中的表 3-9。

① 信源发出的 M 个消息，按其概率递减顺序进行排列，得

$$q(x_1)\geqslant q(x_2)\geqslant q(x_3)\geqslant\cdots\geqslant q(x_M)$$

把消息集$\{x_1, x_2, x_3, \cdots, x_M\}$按其概率大小分解成两个子集，使两个子集的概率之和尽可能相等，把第一个子集编码为“0”，第二个子集编码为“1”，作为代码组的第一个码元；

② 对子集做第二次分解，同样分解成两个子集，并使两个子集的概率之和尽可能接近相等，再把第一个子集编码为“0”，第二个子集编码为“1”，作为代码组的第二个码元；

③ 如此一直进行下去，直到各子集仅含一个消息为止；

④ 将逐次分解过程当中得到的码元排列起来就是各消息的代码。

【例 3.15】 对例 3.14 给出的信源 $\begin{bmatrix}X\\q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7\\0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01\end{bmatrix}$ 进行费诺编码。

（1）将信源消息分成两个子集{x_1，x_2，x_3}和{x_4，x_5，x_6，x_7}，两个子集的概率和分别为 0.2+0.19+0.18=0.57 与 0.17+0.15+0.10+0.01=0.43，赋予第一个子集码元“0”，赋予第二个子集码元“1”；

（2）又将子集分成概率和尽可能接近相等的两个子集，分别赋予第一个子集码元“0”，赋予第二个子集码元“1”；

（3）一直进行下去，直到每个子集仅含一个消息为止。

编码结果列于表 3-9。

表 3-9　费诺编码

消息符号 x_i	消息概率 $q(x_m)$	第一次分解所得码元	第二次分解所得码元	第三次分解所得码元	第四次分解所得码元	码字 $\boldsymbol{c}_m$	码长 n_i
x_1	0.2	0	0			00	2
x_2	0.19			0		010	3
x_3	0.18		1	1		011	3
x_4	0.17	1	0			10	2
x_5	0.15			0		110	3
x_6	0.10		1		0	1110	4
x_7	0.01			1	1	1111	4

从上面得到的码字可以看到，费诺编码法不能保证概率大的消息一定对应码长较短的码字（例如，x_4 的概率比 x_3 小，而它的码字的码长也比 x_3 的码字的码长短），因此费诺码不一定是紧致码，如果每次分得的两个子集都能保证概率相等，则得到的一定是紧致码。

下面计算该编码的编码效率 $\eta=\dfrac{H(X)}{\bar{n}\log D}$。

例 3.14 中已算出信源熵 $H(X)$=2.61（比特/符号）。

平均码长
$$\bar{n}=\sum_{m=1}^{7}n_m q(x_m)$$
$$=2\times(0.2+0.17)+3\times(0.19+0.18+0.15)+4\times(0.10+0.01)$$
$$=2.74$$

则编码效率
$$\eta=\frac{H(X)}{\bar{n}\log D}=\frac{2.61}{2.74\times1}=0.935$$

可见，费诺编码法的编码效率比香农编码法高。

由费诺编码过程可以看出，由于在逐次分解过程中，各子集取码元“0”或“1”是任意的，因此最后得到的代码组并不唯一，但其平均码长 $\bar{n}$ 都是相同的。

3.4.3　霍夫曼编码法

设信源消息数 $M\geqslant2$，记概率分布为 $\begin{bmatrix}X\\q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_M\\q(x_1) & q(x_2) & \cdots & q(x_M)\end{bmatrix}$，存在 D 进制唯一可译码 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M\}$，对应的码长分别为 $\{n_1,n_2,\cdots,n_M\}$，不失一般性，设 $q(x_1)\geqslant q(x_2)\geqslant\cdots\geqslant q(x_M)$，则 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M\}$ 是最佳码必须具备如下两条性质：

① $n_1\leqslant n_2\leqslant\cdots\leqslant n_M$；

② 最后（最长）的 D^* 个码字，它们具有相同的前缀 $\boldsymbol{c}$，唯一的区别是，最后一位码符号不同，可将这 D^* 个最长的码字分别表示为 $\boldsymbol{c}\cdot0$，$\boldsymbol{c}\cdot1$，…，$\boldsymbol{c}\cdot(D^*-1)$

其中
$$D^*\in\{2,3,\cdots,D\}\tag{3-31}$$

且
$$D^*=M[\mathrm{mod}(D-1)]\tag{3-32}$$

上面的两条性质就暗示了一种编码方法，即对于给定信源（信源消息个数为 M）及有 D 个元素的码符号集，设信源概率分布为 $Q=\{q(x_1), q(x_2), \cdots, q(x_M)\}$，对信源进行 D 进制编码，得到码集 $\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M\}$，设码 $\boldsymbol{C}$ 中最长的码字有 D^* 个，这 D^* 个码字的最后一位码符号分别为 0，1，…，D^*-1。

再构成一个新的辅助信源，新信源的概率分布为 $Q^*=\{q(x_1), q(x_2), \cdots, q(x_{M-D^*}), q^*\}$，其中 q^* 是对应原信源中码长最长的 D^* 个码字的概率之和。对新的辅助信源按概率大小重新排序，对其中概率最小的 D 个码字分别赋予后缀 0，1，…，$D-1$。

类似地又可构成新的辅助信源。

下面的定理 3.5 证明对辅助集为最佳的码，对于原始集也是最佳的，反之也成立。

定理 3.5 假定 $\boldsymbol{C}^*=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{M-D}, \boldsymbol{c}\}$ 为最佳码，对应概率分布

$$Q^*=\{q(x_1), q(x_2), \cdots, q(x_{M-D^*}), q^*\}$$

式中，q^* 可记为 $\quad q^*=q(x_{M-D^*+1})+q(x_{M-D^*+2})+\cdots+q(x_M)$

且概率分布满足 $\quad q(x_1)\geqslant q(x_2)\geqslant\cdots\geqslant q(x_{M-D^*})\geqslant q(x_{M-D^*+1})\geqslant\cdots\geqslant q(x_M)$

则对应概率分布为 $Q=\{q(x_1), q(x_2), \cdots, q(x_{M-D^*}), \quad q(x_{M-D^*+1}), \cdots, q(x_M)\}$ 的最佳码是

$$\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{M-D^*}, \boldsymbol{c}\cdot 0, \boldsymbol{c}\cdot 1, \cdots, \boldsymbol{c}\cdot(D^*-1)\}$$

证明：

（1）设已找到 $\boldsymbol{C}^*=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{M-D^*}, \boldsymbol{c}\}$ 是最佳编码，平均码长为 $\bar{n}^*$，又记对应概率分布 Q 的最佳码的平均码长为 $\bar{n}$。

现在由 $\boldsymbol{C}^*$ 来构造 $\boldsymbol{C}$，构造的方法就是，前面的 $M-D^*$ 个码字 $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{M-D^*}$ 保持不变，将最后一个码字 $\boldsymbol{c}$ 分别加上后缀 0, 1, …, D^*-1 构成新的 D^* 个码字，这样构成的码 $\boldsymbol{C}$ 的平均码长为 $\bar{n}^*-nq^*+(n+1)q^*=\bar{n}^*+q^*$，式中，$n$ 是码字 $\boldsymbol{c}$ 的长度，这样构成的码 $\boldsymbol{C}$ 还没有被证明是最佳码，因此有

$$\bar{n}^*+q^*\geq\bar{n} \tag{3-33}$$

（2）又设先找到对应概率分布 Q 的最佳码，即

$$\boldsymbol{C}=\{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{M-D^*}, \boldsymbol{c}\cdot 0, \boldsymbol{c}\cdot 1, \cdots, \boldsymbol{c}\cdot(D^*-1)\}$$

现在由 $\boldsymbol{C}$ 来构造 $\boldsymbol{C}^*$，构造的方法是将 D^*-1 个码字 $\boldsymbol{c}\cdot 0, \boldsymbol{c}\cdot 1, \cdots, \boldsymbol{c}\cdot(D^*-1)$ 中每个码字的最后一位略去，合成一个码字 $\boldsymbol{c}$ 对应概率 q^*，这样构成的码 $\boldsymbol{C}^*$ 的平均码长为 $\bar{n}-1\cdot q^*$，同样，这样构成的码 $\boldsymbol{C}^*$ 还没有被证明是最佳码，因此有

$$\bar{n}-1\cdot q^*\geq\bar{n}^*$$

$$\bar{n}\geq\bar{n}^*+q^* \tag{3-34}$$

综合式（3-33）和式（3-34），得

$$\bar{n}^*+q^*=\bar{n}$$

q^* 与编码后的码长无关，所以若 $\bar{n}^*$ 是最短的，$\bar{n}$ 也一定是最短的。 证毕

【例 3.16】 对例 3.14 给出的信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$ 进行霍夫曼编码，码符号集{0，1，2}，对信源进行 $D=3$ 进制编码。

在此例中，$M=7$，$D=3$。

（1）根据式（3-31）和式（3-32）$\begin{cases} D^*=M[\mathrm{mod}(D-1)] \\ D^*\in\{2,3,\cdots,D\} \end{cases}$，计算出 $M[\mathrm{mod}\,(D-1)]=7[\mathrm{mod}2]=1$；

又根据 $D^*\in\{2，3，\cdots，D\}$，算出 $D^*=3$，即最长的码字数为 3。

将最长的码字分配给概率最小的 3 个消息 x_5, x_6, x_7，并将这三个消息的概率合并为

$$q(x_5)+q(x_6)+q(x_7)=0.15+0.10+0.01=0.26$$

得到含有 5 个消息的新信源，并将之按其概率大小重新排列，即新的辅助信源为

$$\begin{bmatrix} X' \\ q(X') \end{bmatrix}=\begin{bmatrix} x_1' & x_2' & x_3' & x_4' & x_5' \\ 0.26 & 0.20 & 0.19 & 0.18 & 0.17 \end{bmatrix}$$

（2）取 $D=3$，即辅助信源的最长码字数目为 3，把最长的码字分配给概率最小的消息 x_3'，x_4', x_5'，并将这三个消息的概率合并为

$$q(x_3')+q(x_4')+q(x_5')=0.19+0.18+0.17=0.54$$

又得到含有 3 个消息的新信源，并将之按其概率大小重新排列，即新的辅助信源为

$$\begin{bmatrix} X'' \\ q(X'') \end{bmatrix}=\begin{bmatrix} x_1'' & x_2'' & x_3'' \\ 0.54 & 0.26 & 0.20 \end{bmatrix}$$

（3）对上述信源 X, X', X'' 编码如下

$$\begin{bmatrix} X'' \\ q(X'') \\ 编码 \end{bmatrix}=\begin{bmatrix} x_1'' & x_2'' & x_3'' \\ 0.54 & 0.26 & 0.20 \\ 2 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} X' \\ q(X') \\ 编码 \end{bmatrix}=\begin{bmatrix} x_1' & x_2' & x_3' & x_4' & x_5' \\ 0.26 & 0.20 & 0.19 & 0.18 & 0.17 \\ 1 & 0 & 22 & 21 & 20 \end{bmatrix}$$

$$\begin{bmatrix} X \\ q(X) \\ 编码 \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \\ 0 & 22 & 21 & 20 & 12 & 11 & 10 \end{bmatrix}$$

根据定理 3.5，对信源 X″ 为最佳的编码，对 X' 也为最佳，从而对 X 也是最佳的。

下面计算该编码的编码效率 $\eta=\dfrac{H(X)}{\bar{n}\log D}$。

例 3.14 中已算出信源熵 $H(X)=2.61$（比特/符号），计算平均码长

$$\bar{n}=\sum_{m=1}^{7}n_m q(x_m)$$

$$=1\times0.2+2\times0.8=1.8$$

则编码效率

$$\eta=\frac{H(X)}{\bar{n}\log D}=\frac{2.61}{1.8\times\log 3}=\frac{2.61}{1.8\times1.58}=0.92$$

每次这样划分辅助信源很麻烦，可以按照如下步骤编码（先考虑 $D=2$ 的情况），参见例 3.17 中的图 3-7。

① 将信源发出的 M 个消息，按其概率递减顺序进行排列，得

$$q(x_1)\geqslant q(x_2)\geqslant q(x_3)\geqslant\cdots\geqslant q(x_M)$$

② 将概率最小的两个消息分别编码为“1”和“0”，（一般，将概率大的编码为“1”，概率小的编码为“0”），再对这两个消息求概率之和；

③ 将上述概率之和作为一新消息的概率，与余下的消息一起组成一新的信源，再按概率递减顺序重新排列，如果概率之和与原信源的某个概率相等，则把概率之和排在上面，这样可

使合并消息重复编码的次数减少，使短码得到充分利用。与把概率之和排在下面相比，两种方法得到的平均码长 $\overline{n}$ 虽然一样，但后者的方差 $E\{(n-\overline{n})^2\}$ 更大；

④ 如此一直进行下去，直到两个合并消息的概率之和为 1；

⑤ 从最后一步开始，沿编码逆过程取各步骤得到的码符号，如此构成的码符号序列即为对应消息的码字。

【例 3.17】 对例 3.14 给出的信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$ 进行 $D=2$ 进制霍夫曼编码，编码结果如图 3-7 所示。

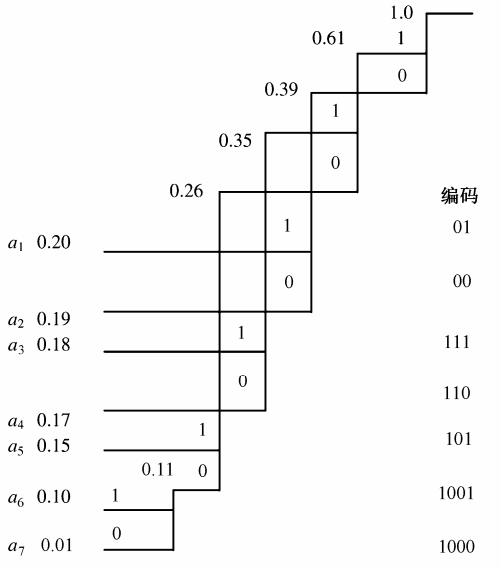


图 3-7 二元霍夫曼编码

下面计算该编码的编码效率 $\eta = \dfrac{H(X)}{\overline{n}\log D}$。

例 3.14 中已算出信源熵 $H(X)=2.61$（比特/符号），计算平均码长

$$\overline{n} = \sum_{m=1}^{7} n_m q(x_m)$$

$$= 2\times0.39 + 3\times0.5 + 4\times0.11 = 2.72$$

则编码效率 $$\eta = \frac{H(X)}{\overline{n}\log D} = \frac{2.61}{2.72\times\log 2} = \frac{2.61}{2.72\times1} = 0.96$$

由霍夫曼算法可看出，得到的码并不是唯一的，因为对于每次选出的两个消息，哪个编码为“1”，哪个编码为“0”是可以任意选取的，由此可得到不同的编码，但平均码长是不变的。

比较例 3.14、例 3.15 和例 3.17，可以看出，对给定的同一信源进行编码，香农编码法的编码效率最低，霍夫曼编码法的编码效率最高，一般情况也是如此。

【例 3.18】 对给定的信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \end{bmatrix}$，用两种方法进行 $D=2$ 进制霍夫曼编码。

方法一：若概率之和与原信源中的某概率相等，将概率之和往上排，如图 3-8 所示。

方法二：若概率之和与原信源中的某概率相等，将概率之和往下排，如图 3-9 所示。

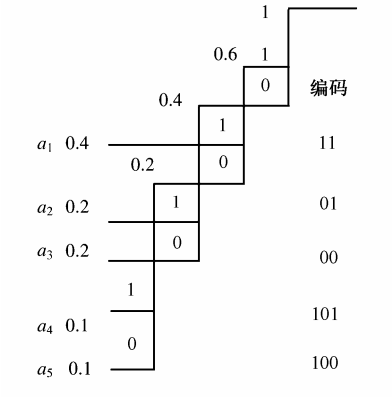


图 3-8　方法一：概率之和往上排

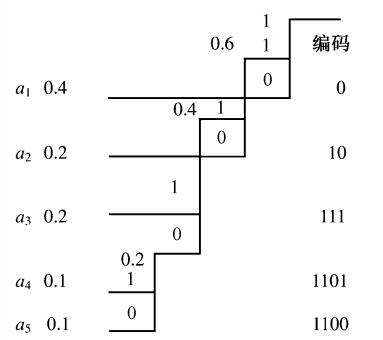


图 3-9　方法二：概率之和往下排

分别计算两种编码法的平均码长 $\overline{n}_1$ 和 $\overline{n}_2$，均方差 $E\{(n-\overline{n}_1)^2\}$ 和 $E\{(n-\overline{n}_2)^2\}$：

$\overline{n}_1 = 2\times 0.8 + 3\times 0.2 = 2.2$

$\overline{n}_2 = 1\times 0.4 + 2\times 0.2 + 3\times 0.2 + 4\times 0.2 = 2.2$

$E\{(n-\overline{n}_1)^2\} = (2-2.2)^2\times 0.8 + (3-2.2)^2\times 0.2 = 0.16$

$E\{(n-\overline{n}_2)^2\} = (1-2.2)^2\times 0.4 + (2-2.2)^2\times 0.2 + (3-2.2)^2\times 0.2 + (4-2.2)^2\times 0.2 = 1.36$

可见，两种编码法平均码长相等，但方法二的均方差较大。

对于 D>2 的 D 进制霍夫曼编码，先根据式（3-31）和式（3-32）$\begin{cases} D^* = M[\mathrm{mod}(D-1)] \\ D^* \in \{2,3,\cdots,D\} \end{cases}$，定出 D^*，即最长的码字数为 D^*，第一次取 D^*个概率合并，以后每次取 D 个概率合并。

【例 3.19】　对给定的信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0.25 & 0.25 & 0.20 & 0.15 & 0.10 & 0.05 \end{bmatrix}$，进行 D=3 进制霍夫曼编码。

对此例有 $M=6, D=3$，先根据 $D^*=M[\ \mathrm{mod}\ (D\text{-}1)]$计算出 D^*=[mod2]=0，再由 $D^*\in\{2,3,\cdots,D\}$，确定 D^*=2。则第一次取两个概率合并，以后每次取三个，如图 3-10 所示。

通过上面的例子可以看出，霍夫曼编码法的指导思想是：概率小的消息赋予较大的码长（重复赋予单码元），概率大的消息赋予较小的码长（重复的次数少）。

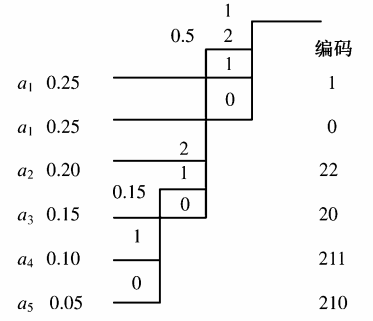


图 3-10　D=3 进制霍夫曼编码

本 章 小 结

信源编码以提高传输效率为主，希望码长尽可能短，本章讨论在不允许失真的前提下对信源的编码，分为两种情况，等长编码和变长编码。等长编码定理和变长编码定理分别给出了这两种情况，在无失真和码长尽可能短这两个约束条件下的平均码长的上界和下界。

1．等长编码定理

记 $H(X)$为单符号信源熵，L 为扩展信源输出序列长度，n 为码字长度，D 为码符号集元素个数，则当满足条件 $\frac{n}{L}>\frac{H(X)+\varepsilon}{\log D}$，$L\to\infty$ 时，可使译码差错 $p_e<\delta$；反之，当 $\frac{n}{L}<\frac{H(X)+\varepsilon}{\log D}$ 时，则不可能实现无差错编码。

2．变长编码定理（Shannon第一定理）

记 $H(X)$为单符号信源熵，L 为扩展信源输出的序列长度，$\bar{n}$ 为信源每个符号所对应的平均码字数，D 为码符号集元素个数，则对信源进行编码，总可以找到一种唯一可译码，使码长 $\bar{n}$ 满足

$$\frac{H(X)}{\log D}\leq \bar{n}<\frac{H(X)}{\log D}+\frac{1}{L}$$

3．平均码长

$$\bar{n}\triangleq\sum_{m=1}^{M}n_m p(c_m)$$

4．克拉夫特不等式

$$\sum_{m=1}^{M}D^{-n_m}\leq 1$$

式中，D 为码符号集元素个数，n_m（$m=1,2,\cdots,M$）是各码字所对应的码长。克拉夫特不等式给出了唯一可译码存在的充要条件，并用于变长编码定理的证明。

本章还介绍了常见的三种变长码的编码方法：香农编码法、费诺编码法和霍夫曼编码法。对于同一信源的编码，三种方法中，以霍夫曼编码的编码效率最高；香农编码法没有太多实用价值，但它在证明变长编码定理时起了重要作用；费诺编码法是遵照变长编码定理（香农第一定理）的指导思想导出的一种编码方法。

思考题与习题

3.1　请问即时码一定是唯一可译码吗？反过来说，唯一可译码一定是即时码吗？

3.2　离散无记忆信源，熵为 $H(X)$，对信源的 L 长序列进行等长编码，码字是长为 n 的 D 进制符号串，问：

（1）满足什么条件时，可实现无失真编码？

（2）L 增大，编码效率 η 也随之增大吗？

3.3　变长编码定理指明，对信源进行变长编码，总可以找到一种唯一可译码，使码长 $\bar{n}$ 满足 $\frac{H(X)}{\log D} \leq \bar{n} < \frac{H(X)}{\log D} + \frac{1}{L}$，试问在 $\bar{n} > \frac{H(X)}{\log D} + \frac{1}{L}$ 时，能否也找到唯一可译码。

3.4　信源的最佳编码使信道码符号等概分布，而且平均码长最短，这种说法对吗？

3.5　对一信源提供 6 种不同的编码方案：码 1 ~ 码 6，见表 3-10。

表 3-10　同一信源的 6 种不同编码

信源消息	消息概率	码 1	码 2	码 3	码 4	码 5	码 6
u_1	1/4	0	001	1	1	00	000
u_2	1/4	10	010	10	01	01	001
u_3	1/8	00	011	100	001	100	011
u_4	1/8	11	100	1000	0001	101	100
u_5	1/8	01	101	10000	00001	110	101
u_6	1/16	001	110	100000	000001	1110	1110
u_7	1/16	111	111	1000000	0000001	1111	1111

（1）说明这些码中哪些是唯一可译码；

（2）说明这些码中哪些是即时码；

（3）对所有唯一可译码求出其平均码长。

3.6　信源有 4 个消息，对其进行二进制编码，问

（1）若信源等概分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$，则每个消息至少需要几位二进制代码？

（2）若信源分布为 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$，则如何编码才能得到紧致码？

3.7　信源符号集 $X = \{0,1,2\}$，一信源含 8 个消息，编码为即时码，若要求码长只取 1，3，5 中之一，应用 Kraft 不等式，分析按上述要求能否构成唯一可译码。

3.8　证明长为 N 的 D 元不等长码至多有 $D(D^N - 1)/(D-1)$ 个码字。

3.9　设信源 S 的 N 次扩展码为 S^N，对其进行霍夫曼编码，码符号集 $B = \{b_1, b_2, \cdots, b_D\}$，编码后得到的码符号 $\{b_1, b_2, \cdots, b_D\}$ 可以看成一个新的信源

$$\begin{bmatrix} B \\ q(B) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & \cdots & b_D \\ p_1 & p_2 & \cdots & p_D \end{bmatrix}$$

证明：当 $N \to \infty$ 时，新信源 B 符号集的概率分布 $p_i \to \frac{1}{D}$（$i = 1, 2, \cdots, D$），即等概分布。

3.10　在例 3.17 中，对信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.2 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$ 进行二进制霍夫曼编码，得到编码效率 $\eta = 0.96$，试问若对上述信源进行二进制等长编码，要求误码率小于 10^{-3}，要达到霍夫曼编码的编码效率 $\eta = 0.96$，估计需要多少个信源符号一起编码才能做到？

3.11　设有一离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ 0.003 & 0.997 \end{bmatrix}$，若对其输出的长为 100 的事件序列中含有两个和更少个 x_1 的序列提供不同的码字，则

（1）在等长编码下，求二元码的最短码长;

（2）求错误概率（误组率）。

3.12　信源符号消息 $X=\{x_1, x_2, \cdots, x_M\}$，信源熵 $H(X)$，若对该信源能找到一个平均码长为 $\overline{n}=\dfrac{H(X)}{\log 3}$ 的三元即时码，证明对每个 $x_i \in X$，其概率满足 $p(x_i)=3^{-n_i}$，式中 n_i 为整数。

3.13　一离散无记忆信源（DMS）$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ 0.9 & 0.1 \end{bmatrix}$，采用下述串长编码法进行编码，见表 3-11。

表 3-11　采用串长法编码

信源输出序列	0 串长度	输出二元码字
1	0	0000
01	1	0001
001	2	0010
0001	3	0011
00001	4	0100
000001	5	0101
0000001	6	0110
00000001	7	0111
00000000	8	1

（1）求信源熵 $H(X)$;

（2）求信源输出序列的平均长度 $\overline{l}$;

（3）求输出二元码字的平均长度 $\overline{n}$;

（4）说明码的唯一可译性。

3.14　离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$,

（1）求 X 的最佳二元码，平均码长及编码效率;

（2）求 $X^{(2)}$ 的最佳二元码，平均码长及编码效率。

3.15　离散无记忆二进制信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，游程计数器检测信源输出序列的连 0 串长度，一个连 0 串是包含在 2 个“1”之间的“0”的个数，例如，从“01000011001000101011001”得到串长“14023102”（串长可大于 10），求该计数器输出的熵和平均串长 $\overline{n}$。

3.16　某一信源有 M 个消息，并且每个消息等概分布，对该信源进行二元霍夫曼编码，问当 $M=2^i$ 和 $M=2^i+1$（i 为正整数）时，每个码字的长度 n^i 等于多少？平均码长 $\overline{n}$ 又等于多少？

3.17 信源分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_3 & x_5 & x_7 \\ 1/3 & 1/3 & 1/4 & 1/12 \end{bmatrix}$，

（1）对该信源进行二元霍夫曼编码;

（2）证明存在两个不同的最佳码长集合，即证明码长集合{1，2，3，3}和{2，2，2，2}都是最佳的。

3.18 对信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 1/5 & 1/6 & 1/6 & 1/10 & 1/10 & 1/10 & 1/12 & 1/12 \end{bmatrix}$，设计

（1）香农编码，$D=2$;

（2）费诺编码，$D=2$;

（3）霍夫曼编码，$D=3$;

（4）霍夫曼编码，$D=4$。

并分别求出平均码长 $\bar{n}$ 和编码效率 η。

3.19 给定信源 S: $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 0.2 & 0.15 & 0.15 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$，对其进行香农编码（$D=2$），

（1）证明编得的码是即时码;

（2）记平均码长为 $\bar{n}$，证明 $H(S)\leq \bar{n}\pi H(S)+1$。

3.20 对一个离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$，构造 D 进制不等长即时码，设平均码长为 $\bar{n}$，记信源熵为 $H(X)$，证明：$\bar{n}=H(X)$ 的充要条件为码的长度集$\{n_1, n_2, \cdots, n_M\}$满足

$$q(x_m)=D^{-n_m}\ (m=1,2,\cdots,M)$$

3.21 离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\ 0.16 & 0.14 & 0.13 & 0.12 & 0.10 & 0.09 & 0.08 & 0.07 & 0.06 & 0.05 \end{bmatrix}$，

（1）求二元霍夫曼编码，计算平均码长 $\bar{n}$ 和编码效率 η;

（2）求三元霍夫曼编码，计算平均码长 $\bar{n}$ 和编码效率 η。

3.22 对信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ 0.2 & 0.15 & 0.15 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$ 进行 D=3 进制霍夫曼编码，要求构造两种不同的码，这两种码的平均码长 $\bar{n}$ 相同，但方差 $\sigma_{\bar{n}}^2 \triangleq E\{(n-\bar{n})^2\}=E\{n^2\}-\bar{n}^2$ 不同，求两种码的 $\bar{n}$ 和 $\sigma_{\bar{n}}^2$，你认为码长 $\bar{n}$ 相同，但方差 $\sigma_{\bar{n}}^2$ 不同的码哪一种对于实用来讲更好。

3.23（1）两个无前缀变长码的级联定义为

$$\boldsymbol{C}=\boldsymbol{C}_1\cdot\boldsymbol{C}_2,\ \text{即}\ \forall c_1\in\boldsymbol{C}_1, \forall c_2\in\boldsymbol{C}_2,\ \ c=c_1\cdot c_2\in\boldsymbol{C}$$

证明：无前缀变长码的级联仍然是无前缀变长码。

（2）考虑以下系统:

设有两个相互独立的信源

$$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_3 & x_5 & x_7 \\ 1/3 & 1/3 & 1/6 & 1/6 \end{bmatrix}\text{和}\begin{bmatrix} Y \\ q(Y) \end{bmatrix}=\begin{bmatrix} y_0 & y_2 & y_4 & y_6 & y_8 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}$$

定义 $z_k=\begin{cases}\frac{1}{2}x_k & k\text{为奇数}\\ \frac{1}{2}y_k & k\text{为偶数}\end{cases}$，$k=0,1,2,3,4,5,6,7,8$。

① $\{z_k\}$的熵为多少？

② 对$\{z_k\}$分别设计 $D=2$，$D=3$ 的霍夫曼编码，比较编码效率η;

③ 对$\{x_k\}$和$\{y_k\}$分别设计 $D=2$，$D=3$ 的霍夫曼编码，将它们级联后得到一个新的无前缀变长码，并将它们的编码效率与②的结果比较;

④ 验证③中的级联码是否仍然满足 Shannon 第一定理（式（3-23））。

3.24 离散无记忆信源 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_0 & x_1\\ 1/4 & 3/4\end{bmatrix}$，

（1）计算该信源的熵;

（2）采用二进制代码传输消息 $x_0=0, x_1=1$，求 $q(0), q(1)$;

（3）对原信源进行二维扩展后，采用费诺编码法，并求编码效率;

（4）对原信源进行三维扩展后，采用霍夫曼编码法，并求编码效率。

3.25 信源 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3 & x_4\\ 0.5 & 0.25 & 0.125 & 0.125\end{bmatrix}$，

（1）对信源进行二进制香农编码;

（2）对信源进行二进制霍夫曼编码;

（3）比较上述两种编码，它们是否完全一致，并说明这种完全一致的一般原理。

3.26 设二元霍夫曼编码为（00，01，10，11）和（0，10，110，111），求出可以编得这样的霍夫曼码的信源的所有概率分布。

第 4 章
离散信道的信道容量

内容提要

第 3 章研究了信源，对信源进行编码，信源编码以提高传输效率为主，在物理信道固定的条件下，总是希望单位时间内传输的信息越多越好，所以信息传输率是衡量通信质量的一个重要指标。但信道对于信息率的容纳并不是无限制的，它不仅与物理信道本身的特性有关，还与信道输入信号的统计特性有关，它有一个极限值，即信道容量，信道容量是有关信道的一个很重要的物理量。本章研究信道，研究在信道中传输的每个符号所携带的信息量，并定义信道容量。

知识要点

信道容量的定义，平均互信息量达到信道容量的充要条件，信道容量的计算，组合信道的容量，数据处理定理。

教学建议

本章定义了信道容量，并给出了信道容量的计算方法，信道容量的计算是件冗长烦琐的事情，文中给出了平均互信息量达到信道容量的充要条件，以及几类特殊信道的信道容量的计算方法，同时简要介绍了几种组合信道的总信道容量。可借助例题掌握这些方法。建议学时数为 5 学时。

4.1 信道容量的定义

根据式（2-35），平均互信息量 $I(X;Y) = H(X) - H(X|Y)$，其中 $H(X)$是信源熵，在图 2-1 所示的简单通信模型中，$H(X)$可看成信道输入方关于发送符号集 X 中的哪个消息的平均不确定性的度量，而 $H(X|Y)$是信道输出方接收到符号集 Y 后仍存在的对于 X 发送哪个消息的平均不确定性的度量，二者之差 $I(X;Y)$就是通信过程中获得的信息量，也就是平均每个码元所携带的信息量。

对于单符号传输的情况，信息传输率

$$R_D = I(X;Y) \tag{4-1}$$

将式（4-1）表示的信息传输率与式（3-4）所定义的信息传输率 $R_D = \dfrac{H(X)}{\overline{n}}$ 相比较，式（4-1）是考虑单符号传输的情况，平均码长 $\overline{n}=1$，而在定义式（3-4）时，没考虑信道的干扰，实际上对于无扰信道 $H(X|Y)=0$，因此有 $I(X;Y)=H(X)-H(X|Y)=H(X)$，所以两个式子的含义不矛盾。

信息传输率是衡量通信质量的一个重要指标，根据式（4-1），研究信息传输率就需要研究 $I(X;Y)$，定理 1.1 指出：当给定信道（即给定信道转移概率 $p(y|x)$）时，平均互信息量 $I(X;Y)$ 是信源输入概率分布 $q(x)$ 的∩型凸函数，所以对于固定信道，总存在某种输入概率分布 $q(x)$，使 $I(X;Y)$达到最大值，定义这个最大值为信道容量，记为 C。

$$C \triangleq \max_{q(x)} I(X;Y) \quad （比特/码符号） \tag{4-2}$$

使 $I(X;Y)$达到信道容量的分布 $q(x)$为最佳分布。

信道容量 C 就是在保证可靠通信的前提下，信道所能容纳的最大信息传输量。

显然，对于固定信道，信道容量 C 是一个定值，对于不同信道，C 值不同，信道容量 C 是信道转移概率 $p(y|x)$的函数。

4.2 离散无记忆信道容量的计算

如果信道输入的是 N 维序列 $\boldsymbol{x}$，其概率分布为 $q(\boldsymbol{x})$，输出的是 N 维序列 $\boldsymbol{y}$，则平均互信息量记为 $I(\boldsymbol{X};\boldsymbol{Y})$，此信道的信道容量 C_N 定义为

$$C_N \triangleq \max_{q(\boldsymbol{x})} I(\boldsymbol{X};\boldsymbol{Y}) \tag{4-3}$$

下面一条定理给出了一维信道和 N 维信道的信道容量之间的关系。

定理 4.1 如果信道是离散无记忆（DMC）的，则 $C_N \leqslant NC$，其中 C 是同一信道传输单符号时的信道容量。

证明：对于 DMC，由定理 2.4 知 $I(\boldsymbol{X};\boldsymbol{Y}) \leq \sum_{i=1}^{N} I(X_i;Y_i)$

$$C_N \triangleq \max_{q(\boldsymbol{x})} I(\boldsymbol{X};\boldsymbol{Y}) \leq \max_{q(\boldsymbol{x})} \sum_{i=1}^{N} I(X_i;Y_i)$$

$$\leq \sum_{i=1}^{N} \max_{q(X_i)} I(X_i;Y_i)$$

$$= \sum_{i=1}^{N} C_i = NC$$

$$\Rightarrow C_N \leqslant NC \qquad (4\text{-}4)$$

证毕

式中，$C_i = \max\limits_{q(X_i)} I(X_i;Y_i)$，这是某时刻 i 允许通过离散无记忆信道的最大信息量，因为序列 $\boldsymbol{x}=x_1 \cdots x_i \cdots x_N$ 中的每个符号都在同一信道中传输，故有 $C_i = C$（$i=1, 2, \cdots, N$）。

（1）若输入的 N 个符号统计独立，满足 $q(\boldsymbol{x}) = \prod\limits_{i=1}^{N} q(x_i)$，即信源离散无记忆，则根据定理 2.3 有

$$I(\boldsymbol{X};\boldsymbol{Y}) \geq \sum_{i=1}^{N} I(X_i;Y_j)$$

（2）对每个 i，输入分布 $q(x_i)$ 可使 $I(X_i; Y_j)$ 达到信道容量 C，则

$$C_N \triangleq \max_{q(\boldsymbol{x})} I(\boldsymbol{X};\boldsymbol{Y})$$

$$\geq \max_{q(x_1)\cdots q(x_N)} \sum_{i=1}^{N} I(X_i;Y_i)$$

$$= \max_{q(x_1)} I(X_1;Y_1) + \max_{q(x_2)} I(X_2;Y_2) + \cdots + \max_{q(x_N)} I(X_N;Y_N)$$

$$= \sum_{i=1}^{N} C_i = NC$$

即

$$C_N \geqslant NC \qquad (4\text{-}5)$$

综合式（4-4）和式（4-5），可知在信源和信道都离散无记忆的情况下，有 $C_N = NC$，即定理中等号成立，这时 N 长序列的传输问题可归结为单符号传输问题。

4.2.1 达到信道容量的充要条件

下面的定理给出了使平均互信息量 $I(X; Y)$达到信道容量 C 的充要条件。

定理 4.2 使平均互信息量 $I(X;Y)$达到信道容量 C 的充要条件是，信道输入概率分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ q(x_1) & q(x_2) & \cdots & q(x_M) \end{bmatrix}$，简记为 $\boldsymbol{q}(X) = \{q(x_1), q(x_2), \cdots, q(x_M)\}$，满足

$$\begin{cases} I(x_i;Y) = C & \text{若 } q(x_i) > 0 \\ I(x_i;Y) \leqslant C & \text{若 } q(x_i) = 0 \end{cases} \quad i = 1, 2, \cdots, M \qquad (4\text{-}6)$$

这条定理只给出了使平均互信息量 $I(X; Y)$达到信道容量 C 的充要条件，并没有给出求 C 及分布 $q(x)$的显式，因此除了一些特殊情况可利用这一定理求解外，一般情况下要利用它来求解是困难的，下面举例说明它的应用。

【例 4.1】 信道输入符号集 $X = \{x_1, x_2, x_3\}$，输出符号集 $Y = \{y_1, y_2\}$，信道转移概率矩阵 $\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{bmatrix}$，求信道容量和最佳分布。

根据矩阵中元素 $p(y|x)$的分布，可以设想最佳分布为 $q(x_1)=0.5$，$q(x_2)=0$，$q(x_3)=0.5$，这样信道的输入和输出就构成了一一对应的关系，如图 4-1 所示，接收方收到输出符号集 Y 中的某个元素后，对发送方发送的是输入符号集 X 中的哪个元素是完全确定的，即 $H(X\mid Y)=0$。

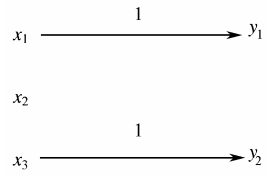


图 4-1　取 $q(x_2)=0$ 时信道输入和输出的对应关系

若取 $q(x_2)\neq 0$，则接收方收到 $Y=\{y_1,y_2\}$中的某个符号后，关于发送的是 $X=\{x_1,x_2,x_3\}$中的哪个符号就会增加不确定性，即 $H(X\mid Y)\neq 0$，从平均互信息量的表达式 $I(X;Y)=H(X)-H(X\mid Y)$ 可看出，由于 $H(X\mid Y)$的非负性，这样得到的平均互信息量就会减少，所以直观地看，按设想的分布得到的 $I(X;Y)$应取最大值。

下面还要验证一下上述分布是否满足定理 4.2 所要求的充要条件，先算出

$$\begin{cases}\omega(y_1)=\sum\limits_i q(x_i)p(y_1|x_i)=0.5\times 1=0.5\\ \omega(y_2)=\sum\limits_i q(x_i)p(y_2|x_i)=0.5\times 1=0.5\end{cases}$$

$$\begin{cases}I(x_1;Y)=\sum\limits_j p(y_j|x_1)\log\dfrac{p(y_j|x_1)}{\omega(y_j)}=1\times\log\dfrac{1}{1/2}+0\times\log\dfrac{0}{1/2}=\log 2\\ I(x_2;Y)=\sum\limits_j p(y_j|x_2)\log\dfrac{p(j|x_2)}{\omega(y_j)}=\dfrac{1}{2}\times\log\dfrac{1/2}{1/2}+\dfrac{1}{2}\times\log\dfrac{1/2}{1/2}=0\\ I(x_3;Y)=\sum\limits_j p(y_j|x_3)\log\dfrac{p(j|x_3)}{\omega(y_j)}=0\times\log\dfrac{0}{1/2}+1\times\log\dfrac{1}{1/2}=\log 2\end{cases}$$

显然满足定理 4.2 的充要条件为 $\begin{cases}I(x_1;Y)=I(x_3;Y)=\log 2 & q(x_1)\neq 0\quad q(x_3)\neq 0\\ I(x_2;Y)=0 & q(x_2)=0\end{cases}$

故信道容量 $C=\log 2=1$（比特/符号）。

下面举例介绍三类特殊信道：无噪确定信道、有噪无损信道和无噪无损信道，这些信道的输入 X 和输出 Y 之间有着确定的关系。

【例 4.2】　无噪确定信道。

无噪确定信道的输入符号集元素个数小于输出符号集的元素个数，信道的一个输入对应多个互不交叉的输出，如图 4-2 所示，信道输入符号集 $X=\{x_1,x_2,x_3\}$，输出符号集 $Y=\{y_1,y_2,y_3,y_4,y_5\}$，其信道转移概率矩阵记为 $\boldsymbol{P}=\begin{bmatrix}1&0&0&0&0&0\\0&1/2&1/2&0&0&0\\0&0&0&1/6&2/6&3/6\end{bmatrix}$，

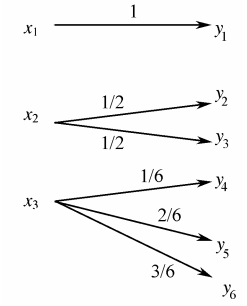


图 4-2　无噪确定信道

计算该信道的信道容量。

（1）先考察平均互信息量 $I(X;Y)=H(X)-H(X\mid Y)$，式中的 $H(X\mid Y)$是收到 $Y=\{y_1,y_2,y_3,y_4,y_5,y_6\}$中的某个符号后，关于发送的是 $X=\{x_1,x_2,x_3\}$中的哪个符号仍然保留的疑义度，在无噪信道条件下，$H(X\mid Y)=0$，则平均互信息量 $I(X;Y)=H(X)$。

（2）根据定义计算信道容量 C

$$C\triangleq\max_{q(x)}I(X;Y)=\max_{q(x)}H(X)$$

从上式可看出，求信道容量 C 的问题转化为寻找某种分布 $q(x)$ 使信源熵 $H(X)$达到最大，由极大离散熵定理可知，在信源消息等概分布 $q(x_1)=q(x_2)=q(x_3)=\frac{1}{3}$时，熵值达到最大，即有

$$C=\max_{q(x)}H(X)=\log 3$$

（3）根据平均互信息量 $I(X;Y)$达到信道容量的充要条件式（4-6）对 C 进行验证。

先根据 $\omega(y_j)=\sum_{i=1}^{3}q(x_i)p(y_j|x_i)$ 计算出$\omega(y_j)$，$j=1,2,3,4,5,6$。

$$\left.\begin{aligned}
\omega(y_1)&=\sum_{i=1}^{3}q(x_i)p(y_1|x_i)=\frac{1}{3}\times 1=\frac{1}{3}\\
\omega(y_2)&=\sum_{i=1}^{3}q(x_i)p(y_2|x_i)=\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}\\
\omega(y_3)&=\sum_{i=1}^{3}q(x_i)p(y_3|x_i)=\frac{1}{3}\times\frac{1}{2}=\frac{1}{6}\\
\omega(y_4)&=\sum_{i=1}^{3}q(x_i)p(y_4|x_i)=\frac{1}{3}\times\frac{1}{6}=\frac{1}{18}\\
\omega(y_5)&=\sum_{i=1}^{3}q(x_i)p(y_5|x_i)=\frac{1}{3}\times\frac{2}{6}=\frac{1}{9}\\
\omega(y_6)&=\sum_{i=1}^{3}q(x_i)p(y_6|x_i)=\frac{1}{3}\times\frac{3}{6}=\frac{1}{6}
\end{aligned}\right\}\text{满足}\sum_{j=1}^{6}\omega(y_j)=1$$

再计算出

$$I(x_1;Y)=\sum_{j=1}^{6}p(y_j|x_1)\log\frac{p(y_j|x_1)}{\omega(y_j)}=1\times\log\frac{1}{1/3}=\log 3$$

$$I(x_2;Y)=\sum_{j=1}^{6}p(y_j|x_2)\log\frac{p(y_j|x_2)}{\omega(y_j)}=\frac{1}{2}\times\log\frac{1/2}{1/6}+\frac{1}{2}\log\frac{1/2}{1/6}=\log 3$$

$$I(x_3;X)=\sum_{j=1}^{6}p(y_j|x_3)\log\frac{p(y_j|x_3)}{\omega(y_j)}=\frac{1}{6}\times\log\frac{1/6}{1/18}+\frac{2}{6}\log\frac{2/6}{1/9}+\frac{3}{6}\log\frac{3/6}{1/6}=\log 3$$

上面三式满足平均互信息量达到信道容量 C 的充要条件，故 $C=\log 3$。

【例 4.3】 有噪无损信道。

有噪无损信道的输入符号集元素个数大于输出符号集的元素个数，信道的一个输出对应多个互不交叉的输入，这时输入符号以确定概率 1 指向某个输出符号，如图 4-3 所示，信道输入符号集 $X=\{x_1, x_2, x_3, x_4, x_5\}$，输出符号集 $Y=\{y_1, y_2\}$，其信道转移概率矩阵记为 $\boldsymbol{P}=\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\0&1\end{bmatrix}$，计算该信道的信道容量。

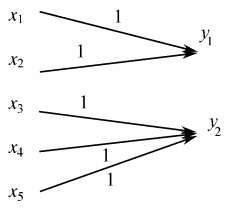


图 4-3 有噪无损信道

（1）先考察平均互信息量 $I(X;Y)=H(Y)-H(Y|X)$，式中的 $H(Y|X)$反映了发送方发出 X

$= \{x_1, x_2, x_3, x_4, x_5\}$中的某个符号后，由于信道干扰而使发送符号错位的一种发散程度，对于无损信道 $H(Y \mid X) = 0$，则平均互信息量 $I(X; Y) = H(Y)$。

（2）根据定义计算信道容量 C

$$C \triangleq \max_{q(x)} I(X;Y) = \max_{q(x)} H(Y)$$

因为 $\omega(y) = \sum_x p(y|x)q(x)$，由于信道转移概率 $p(y|x)$是确定的，求 $q(x)$ 最佳分布的问题就转化为求$\omega(y)$最佳。由极大离散熵定理知道，在$\omega(y)$等概分布 $\omega(y_1) = \omega(y_2) = \frac{1}{2}$ 时，$H(Y)$值达到最大，则

$$C = \max_{\omega(y)} H(Y) = \log 2$$

（3）根据平均互信息量 $I(X; Y)$达到信道容量的充要条件式（4-6）对 C 进行验证。

计算出

$$I(x_1;Y) = \sum_{j=1}^{2} p(y_j|x_1) \log \frac{p(y_j|x_1)}{\omega(y_j)} = 1 \times \log \frac{1}{1/2} = \log 2$$

$$I(x_2;Y) = \sum_{j=1}^{2} p(y_j|x_2) \log \frac{p(y_j|x_2)}{\omega(y_j)} = 1 \times \log \frac{1}{1/2} = \log 2$$

$$I(x_3;Y) = \sum_{j=1}^{2} p(y_j|x_3) \log \frac{p(y_j|x_3)}{\omega(y_j)} = 1 \times \log \frac{1}{1/2} = \log 2$$

$$I(x_4;Y) = \sum_{j=1}^{2} p(y_j|x_4) \log \frac{p(y_j|x_4)}{\omega(y_j)} = 1 \times \log \frac{1}{1/2} = \log 2$$

$$I(x_5;Y) = \sum_{j=1}^{2} p(y_j|x_5) \log \frac{p(y_j|x_5)}{\omega(y_j)} = 1 \times \log \frac{1}{1/2} = \log 2$$

上面 5 个公式满足平均互信息量达到信道容量 C 的充要条件，故 $C = \log 2$。

【例 4.4】 无噪无损信道。

无噪无损信道的输入符号集的元素个数等于输出符号集的元素个数，且信道的输入符号以确定概率 1 指向某个固定的输出符号，如图 4-4 所示，信道输入符号集 $X = \{x_1, x_2, x_3, x_4, x_5\}$，输出符号集 $Y = \{y_1, y_2, y_3, y_4, y_5\}$，其信道转移概率矩阵记为 $\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$，计算该信道的信道容量。

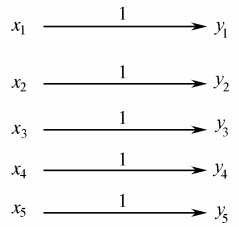


图 4-4 无噪无损信道

信道容量 $C \triangleq \max_{q(x)} I(X;Y) = \max_{q(x)} H(X) = \log 5$。

对于一般的离散信道，要求得信道容量 C 是很困难的，要进行多次试算。更一般地，可以采用迭代算法来计算信道容量，这在不少教科书中都有介绍。

下面介绍几类比较特殊的信道的信道容量的算法。

4.2.2 几类特殊的信道

1. 准对称信道

定义 4.1 如果信道转移概率矩阵 $\boldsymbol{P}$ 中，每一行元素都是另一行相同元素的不同排列，则称该信道关于行（输入）对称。

定义 4.2 如果信道转移概率矩阵 $\boldsymbol{P}$ 中，每一列元素都是另一列相同元素的不同排列，则称该信道关于列（输出）对称。

定义 4.3 如果信道转移概率矩阵 $\boldsymbol{P}$ 可按输出符号集 Y 分成几个子集（子矩阵），而每一子集关于行、列都对称，则称此信道为准对称信道。

若划分的子集只有一个，则该信道关于行（输入）、列（输出）都是对称的，这类信道称为对称信道，对称信道是准对称信道的特例。

【例 4.5】

$\boldsymbol{P}=\begin{bmatrix}0.1 & 0.2 & 0.7\\ 0.2 & 0.7 & 0.1\end{bmatrix}$ 是行对称信道；

$\boldsymbol{P}=\begin{bmatrix}1 & 0\\ 0.5 & 0.5\\ 0 & 1\end{bmatrix}$ 是列对称信道；

$\boldsymbol{P}=\begin{bmatrix}0.3 & 0.2 & 0.5\\ 0.5 & 0.2 & 0.3\end{bmatrix}$ 是准对称信道，可将它划分为 $\begin{bmatrix}0.3 & 0.5\\ 0.5 & 0.3\end{bmatrix}$ 和 $\begin{bmatrix}0.2\\ 0.2\end{bmatrix}$ 两个对称子集；

$\boldsymbol{P}=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}$ 和 $\boldsymbol{P}=\begin{bmatrix}0.3 & 0.7\\ 0.7 & 0.3\end{bmatrix}$ 是对称信道。

定理 4.3 实现 DMC 准对称信道的信道容量的分布为等概分布。

证明：设信道转移概率矩阵 $\boldsymbol{P}$ 中有 S 个对称子集 Y_s（$s=1, 2, \cdots, S$），若信源含 K 个消息，等概分布，即 $q(x_k)=\dfrac{1}{K}$，$k=1,2,\cdots,K$，则

$$
\begin{aligned}
I(x_k,Y)&=\sum_j p(y_j|x_k)\log\frac{p(y_j|x_k)}{\omega(y_j)}\\
&=\sum_j p(y_j|x_k)\log\frac{p(y_j|x_k)}{\sum_i q(x_i)p(y_j|x_i)}\\
&=\sum_j p(y_j|x_k)\log\frac{p(y_j|x_k)}{\frac{1}{K}\sum_i p(y_j|x_i)}\\
&=\sum_{s=1}^{S}\left(\sum_{j\in Y_s} p(y_j|x_k)\log\frac{p(y_j|x_k)}{\frac{1}{K}\sum_i p(y_j|x_i)}\right)
\end{aligned}
$$

因为每个 Y_s 中的行、列对称，对任何 k 值，上式括弧中的和值 $\sum_{j\in Y_s} p(y_j|x_k)\log\dfrac{p(y_j|x_k)}{\dfrac{1}{K}\sum_i p(y_j|x_i)}$ 都相等，与标号 k 无关，故 $I(x_k; Y)$ 是常值，根据定理 4.2，知平均互信息量 $I(X; Y)$达到信道容量 C。 证毕

【例 4.6】 信道输入符号集 $X=\{x_1, x_2\}$，输出符号集 $Y=\{y_1, y_2, y_3, y_4\}$，给定信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1/4 & 1/2 & 1/8 & 1/8\\ 1/8 & 1/2 & 1/4 & 1/8\end{bmatrix}$，求该信道的信道容量 C。

这是一个准对称信道，根据定理 4.3，当 X 等概分布，即 $q(x_1)=q(x_2)=\dfrac{1}{2}$ 时，信道容量

$$C=I(X;Y)\Big|_{q(x_1)=q(x_2)=\frac{1}{2}}$$

平均互信息量 $I(X; Y)=H(Y)-H(Y|X)$

$$=-\sum_{j=1}^{4}\omega(y_j)\log\omega(y_j)+\sum_{i=1}^{2}\sum_{j=1}^{4}q(x_i)p(y_j|x_i)\log p(y_j|x_i) \qquad (4\text{-}7)$$

由 $\omega(y_j)=\sum_{i=1}^{2}q(x_i)p(y_j|x_i)=\dfrac{1}{2}\sum_{i=1}^{2}p(y_j\mid x_i)$，先算出

$$\begin{cases}\omega(y_1)=\dfrac{1}{2}\sum_{i=1}^{2}p(y_1\mid x_i)=\dfrac{1}{2}\times(\dfrac{1}{4}+\dfrac{1}{8})=\dfrac{3}{16}\\ \omega(y_2)=\dfrac{1}{2}\sum_{i=1}^{2}p(y_2\mid x_i)=\dfrac{1}{2}\times(\dfrac{1}{2}+\dfrac{1}{2})=\dfrac{1}{2}\\ \omega(y_3)=\dfrac{1}{2}\sum_{i=1}^{2}p(y_3\mid x_i)=\dfrac{1}{2}\times(\dfrac{1}{8}+\dfrac{1}{4})=\dfrac{3}{16}\\ \omega(y_4)=\dfrac{1}{2}\sum_{i=1}^{2}p(y_4\mid x_i)=\dfrac{1}{2}\times(\dfrac{1}{8}+\dfrac{1}{8})=\dfrac{1}{8}\end{cases} \qquad (4\text{-}8)$$

将式（4-8）和 $q(x_1)=q(x_2)=\dfrac{1}{2}$ 代入式（4-7），可算得信道容量

$$\begin{aligned}C&=I(X;Y)\Big|_{q(x_1)=q(x_2)=\frac{1}{2}}\\ &=-\sum_{j=1}^{4}\omega(y_j)\log\omega(y_j)+\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{4}p(y_j|x_i)\log p(y_j|x_i)\\ &=-2\times\frac{3}{16}\log\frac{3}{16}-\frac{1}{2}\log\frac{1}{2}-\frac{1}{8}\log\frac{1}{8}+\frac{1}{2}\left(4\times\frac{1}{8}\log\frac{1}{8}+2\times\frac{1}{2}\log\frac{1}{2}+2\times\frac{1}{4}\log\frac{1}{4}\right)\\ &=0.0325\text{（比特/符号）}\end{aligned}$$

【例 4.7】 信道输入符号集 $X=\{x_1, x_2, \cdots, x_K\}$，输出符号集 $Y=\{y_1, y_2, \cdots, y_K\}$，给定信道转移概率 $p(y_j|x_i)=\begin{cases}1-\varepsilon & i=j\\ \dfrac{\varepsilon}{K-1} & i\neq j\end{cases}(i,j=1,2,\cdots,K)$，写成矩阵形式为 $\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \dfrac{\varepsilon}{K-1} & \cdots & \dfrac{\varepsilon}{K-1}\\ \dfrac{\varepsilon}{K-1} & 1-\varepsilon & \cdots & \dfrac{\varepsilon}{K-1}\\ \vdots & \vdots & \cdots & \vdots\\ \dfrac{\varepsilon}{K-1} & \dfrac{\varepsilon}{K-1} & \cdots & 1-\varepsilon\end{bmatrix}$，计

算该信道的信道容量 C。

这是一个对称矩阵，根据定理 4.3，当 X 等概分布，即 $q(x_i)=\frac{1}{K}, i=1,2,\cdots,K$ 时，信道容量

$$C=I(X;Y)\Big|_{q(x_i)=\frac{1}{K}}$$

（1）先算出

$$\omega(y_j)=\sum_{i=1}^{K}q(x_i)p(y_j|x_i)=\frac{1}{K}\sum_{i=1}^{K}p(y_j|x_i)$$

$$=\frac{1}{K}\left[(1-\varepsilon)+(K-1)\frac{\varepsilon}{K-1}\right]=\frac{1}{K}\qquad j=1,2,\cdots,K$$

（2）再计算 $C=I(X;Y)\Big|_{\substack{q(x_i)=1/K\\ \omega(y_j)=1/K}}$

$$=\left[-\sum_{j=1}^{K}\omega_j\log\omega_j+\sum_{i=1}^{K}q(x_i)\sum_{j=1}^{K}p(y_j|x_i)\log p(y_j|x_i)\right]\Bigg|_{\substack{q(x_i)=1/K\\ \omega(y_j)=1/K}}$$

$$=\log K+\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{K}p(y_j|x_i)\log p(y_j|x_i)$$

$$=\log K+\frac{1}{K}\times K\left[(1-\varepsilon)\log(1-\varepsilon)+(K-1)\frac{\varepsilon}{K-1}\log\frac{\varepsilon}{K-1}\right]$$

$$=\log K-\varepsilon\log(K-1)-H_2(\varepsilon)$$

式中，

$$H_2(\varepsilon)=-\varepsilon\log\varepsilon-(1-\varepsilon)\log(1-\varepsilon)$$

（3）讨论

当 $K=2$ 时就得到二元对称信道 BSC，有

$$\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}$$

其信道容量 $C=\log 2-H_2(\varepsilon)=1-H_2(\varepsilon)$，如图 4-5 所示。从图中可看出，$k=2$ 时，C–ε曲线关于$\varepsilon=0.5$ 对称。

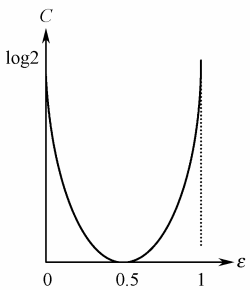


图 4-5　C-ε 曲线

讨论几个特殊点：

① $\varepsilon=0$ 时，信道的输入符号和输出符号是一一对应的关系，如图 4-6 所示，在这种情况下，信道容量 $C=\log 2$，达到最大值。

② $\varepsilon=0.5$ 时，信道的不确定性最大，如图 4-7 所示，在这种情况下，信道容量 $C=0$，这是一种最差的信道。

③ $\varepsilon=1$ 时，这是一种强噪信道，但也是一种确定信道，如图 4-8 所示，在这种情况下，可将判决取反，收到 x_1 判为 y_2，收到 x_2 判为 y_1，也能达到信道容量的最大值 $C=\log 2$。

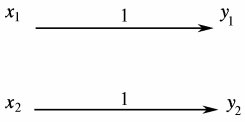


图 4-6　$\varepsilon=0$ 时的信道

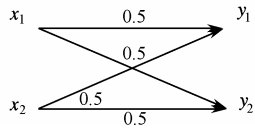


图 4-7　$\varepsilon=0.5$ 时的信道

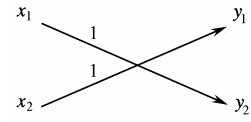


图 4-8　$\varepsilon=1$ 时的信道

2. 信源只含两个消息

用一个例子来说明这种信道容量的计算方法。

【例 4.8】 信道输入符号集 $X=\{x_1, x_2\}$，输出符号集 $Y=\{y_1, y_2, y_3\}$，给定信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-q & q & 0\\ 0 & q & 1-q\end{bmatrix}$，求信道容量 C。

设使平均互信息量达到信道容量的信源分布为 $q(x_1)=\alpha$，$q(x_2)=1-\alpha$。

由
$$\omega(y_j)=\sum_{i=1}^{2}q(x_i)p(y_j|x_i)\qquad j=1,2,3$$

可算出
$$\begin{cases}\omega(y_1)=\sum_{i=1}^{2}q(x_i)p(y_1|x_i)=\alpha(1-q)\\ \omega(y_2)=\sum_{i=1}^{2}q(x_i)p(y_2|x_i)=\alpha q+(1-\alpha)q=q\\ \omega(y_3)=\sum_{i=1}^{2}q(x_i)p(y_3|x_i)=(1-\alpha)(1-q)\end{cases}$$

平均互信息量

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=-\sum_{j=1}^{3}\omega(y_j)\log\omega(y_j)+\sum_{i=1}^{2}\sum_{j=1}^{3}q(x_i)p(y_j|x_i)\log p(y_j|x_i)\\&=-\alpha(1-q)\log\alpha(1-q)-q\log q-(1-\alpha)(1-q)\log(1-\alpha)(1-q)+\\&\quad\alpha[(1-q)\log(1-q)+q\log q]+(1-\alpha)[q\log q+(1-q)\log(1-q)]\\&=-(1-q)[\alpha\log\alpha+(1-\alpha)\log(1-\alpha)]\end{aligned}$$

根据定义，求 C 的问题就转化为α为何值时，$I(X;Y)$ 达到最大值。

令 $\dfrac{\partial I(X;Y)}{\partial\alpha}=0$，有 $\alpha=0.5$。

则信道容量
$$C=I(X;Y)\big|_{\alpha=0.5}=1-q$$

3. 信道转移概率矩阵为非奇异方阵

设信源含 K 个消息$[x_1, x_2, \cdots, x_K]$，对应概率分布

$$\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & \cdots & x_K\\ q(x_1) & q(x_2) & \cdots & q(x_K)\end{bmatrix}$$

若信道转移概率矩阵 $\boldsymbol{P}$ 为非奇异方阵，即满足

$$\begin{cases}[p(y_j|x_i)]=\boldsymbol{P}\text{(方阵)}\\ \text{行列式}\ \left|[p(y_j|x_i)]\right|\neq 0\end{cases}$$

则逆方阵$[p(y_j|x_i)]^{-1}=\boldsymbol{P}^{-1}$存在。

求信道容量 C 的问题就是在以下约束条件下，求平均互信息量 $I(X;Y)$的极值。

$$\begin{cases}\sum_{i=1}^{K}q(x_i)=1 & \text{(4-9)}\\ q(x_i)\geq 0\qquad i=1,2,\cdots,K & \text{(4-10)}\end{cases}$$

先不考虑第二个约束条件式（4-10），用拉格朗日乘因子法求函数极值。

构造函数：

$$\Phi = I(X;Y) - \lambda\sum_{i=1}^{K} q(x_i)$$

$$= -\sum_{j=1}^{K} \omega(y_j)\log\omega(y_j) + \sum_{i=1}^{K}\sum_{j=1}^{K} q(x_i)p(y_j|x_i)\log p(y_j|x_i) - \lambda\sum_{i=1}^{K} q(x_i)$$

式中，λ为拉格朗日乘因子（待定常数）。

令 $\frac{\partial\Phi}{\partial q(x_i)} = 0$，为计算方便，取 e 为对数底数，注意到

$$\omega(y_j) = \sum_{i=1}^{K} q(x_i)p(y_j|x_i)$$

则

$$\frac{\partial\omega(y_j)}{\partial q(x_i)} = p(y_j|x_i)$$

可得 $\frac{\partial\Phi}{\partial q(x_i)} = -\sum_{j=1}^{K} p(y_j|x_i)\ln\omega(y_j) - \sum_{j=1}^{K} p(y_j|x_i) + \sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i) - \lambda = 0$

$$-\sum_{j=1}^{K} p(y_j|x_i)\ln\omega(y_j) + \sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i) = \lambda + 1 \qquad (4\text{-}11)$$

将式（4-11）两边同时乘以 $q(x_i)$，并对 i 求和，得

$$-\sum_{i=1}^{K} q(x_i)\sum_{j=1}^{K} p(y_j|x_i)\ln\omega(y_j) + \sum_{i=1}^{K} q(x_i)\sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i) = \sum_{i=1}^{K} q(x_i)(\lambda+1)$$

$$-\sum_{j=1}^{K} \omega(y_j)\ln\omega(y_j) + \sum_{i=1}^{K} q(x_i)\sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i) = \lambda + 1 \qquad (4\text{-}12)$$

式（4-12）左边恰好就是平均互信息量 $I(X;Y)$的表达式，即有

$$I(X;Y) = \lambda + 1 \overset{\text{记为}}{=\!=} C \quad \text{（待定常数）}$$

这时 $I(X;Y)$ 已达到最大值 C（信道容量），只要进一步求出待定常数 C 及对应的分布 $q(x_i)$（i =1, 2, …, K）即可。

$p(y_j \mid x_i)$是矩阵 $\boldsymbol{P}$的第 i 行第 j 列元素，记 $p^{-1}(y_j \mid x_k)$ 是矩阵 $\boldsymbol{P}^{-1}$ 的第 k 行第 i 列元素，因为矩阵 $\boldsymbol{P}$和矩阵 $\boldsymbol{P}^{-1}$ 互逆，满足 $\boldsymbol{P}\boldsymbol{P}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{P} = \boldsymbol{E}$，$\boldsymbol{E}$为单位矩阵，故有

$$\sum_{i=1}^{K} p(y_j|x_i)\cdot p^{-1}(y_i|x_k) = \delta_{jk} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases} \qquad (4\text{-}13)$$

另外，

$$\sum_{i=1}^{K} p^{-1}(y_i|x_k) = \sum_{j=1}^{K} p(y_j|x_i)\sum_{i=1}^{K} p^{-1}(y_i|x_k)$$

$$= \sum_{j=1}^{K}\left[\sum_{i=1}^{K} p(y_j|x_i)\cdot p^{-1}(y_i|x_k)\right]$$

$$= \sum_{j=1}^{K} \delta_{jk} = 1$$

即

$$\sum_{i=1}^{K} p^{-1}(y_i|x_k) = 1 \qquad (4\text{-}14)$$

利用式（4-14）的结果，用 $p^{-1}(y_i \mid x_k)$乘以式（4-11）两边，并对 i 求和，有

$$-\sum_{i=1}^{K} p^{-1}(y_i|x_k)\sum_{j=1}^{K} p(y_j|x_i)\ln\omega(y_j)+\sum_{i=1}^{K} p^{-1}(y_i|x_k)\sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i)=C\sum_{i=1}^{K} p^{-1}(y_i|x_k)$$

$$-\sum_{j=1}^{K}\left[\ln\omega(y_j)\sum_{i=1}^{K} p^{-1}(y_i|x_k)p(y_j|x_i)\right]+\sum_{i=1}^{K}\sum_{j=1}^{K} p^{-1}(y_i|x_k)p(y_j|x_i)\ln p(y_j|x_i)=C \tag{4-15}$$

将式（4-13）代入式（4-15），得

$$-\sum_{j=1}^{K}\delta_{jk}\ln\omega(y_j)+\sum_{i=1}^{K}\sum_{j=1}^{K} p^{-1}(y_i|x_k)p(y_j|x_i)\ln p(y_j|x_i)=C$$

$$-\ln\omega(y_k)+\sum_{i=1}^{K}\sum_{j=1}^{K} p^{-1}(y_i|x_k)p(y_j|x_i)\ln p(y_j|x_i)=C$$

$$\omega(y_k)=\exp\left\{-C+\sum_{i=1}^{K} p^{-1}(y_i|x_k)\left[\sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i)\right]\right\}$$

$$\omega(y_k)=\exp\left\{-C-\sum_{i=1}^{K} p^{-1}(y_i|x_k)H(Y|x_i)\right\} \tag{4-16}$$

式（4-16）中

$$H(Y|x_i)=-\sum_{j=1}^{K} p(y_j|x_i)\ln p(y_j|x_i) \tag{4-17}$$

在式（4-16）中对 k 求和，得

$$1=\sum_{k=1}^{K}\exp\left\{-C-\sum_{i=1}^{K} p^{-1}(y_i|x_k)H(Y|x_i)\right\}$$

$$C=\ln\sum_{k=1}^{K}\exp\left\{-\sum_{i=1}^{K} p^{-1}(y_i|x_k)H(Y|x_i)\right\} \tag{4-18}$$

式（4-18）就是信道容量 C 的计算公式，等式右边都是已知量。

我们在推导信道容量 C 的计算公式时，并没有考虑式（4-23）所示的约束条件 $q(x_i)\geqslant 0$，$i=1, 2, \cdots, K$，所以还应计算 C 所对应的分布 $q(x_i)$，看是否满足 $q(x_i)\geqslant 0$。

$q(x_i)$ 的计算公式可如下考虑，已知

$$\omega(y_j)=\sum_{i=1}^{K} q(x_i)p(y_j|x_i)\qquad j=1,2,\cdots,K \tag{4-19}$$

将式（4-19）的 K 个式子写成矩阵形式，有

$$\left[\omega(y_1)\ \cdots\ \omega(y_K)\right]=\left[q(x_1)\ \cdots\ q(x_K)\right]\begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_K|x_1)\\ p(y_1|x_2) & p(y_2|x_2) & \cdots & p(y_K|x_2)\\ \vdots & \vdots & \vdots & \vdots\\ p(y_1|x_K) & p(y_2|x_K) & \cdots & p(y_K|x_K)\end{bmatrix}$$

简记为

$$\boldsymbol{\Omega = Q\ P} \tag{4-20}$$

用 $\boldsymbol{P}$ 的逆矩阵 $\boldsymbol{P}^{-1}$ 乘以式（4-20）两边，得

$$\boldsymbol{\Omega P^{-1}=QPP^{-1}}$$

$$\boldsymbol{\Omega P^{-1}=Q}$$

即

$$\left[q(x_1) \quad \cdots \quad q(x_K)\right]=\left[\omega(y_1) \quad \cdots \quad \omega(y_K)\right]\cdot\begin{bmatrix} p^{-1}(y_1|x_1) & p^{-1}(y_2|x_1) & \cdots & p^{-1}(y_K|x_1) \\ p^{-1}(y_1|x_2) & p^{-1}(y_2|x_2) & \cdots & p^{-1}(y_K|x_2) \\ \vdots & \vdots & \vdots & \vdots \\ p^{-1}(y_1|x_K) & p^{-1}(y_2|x_K) & \cdots & p^{-1}(y_K|x_K) \end{bmatrix}$$

式中，

$$q(x_i)=\sum_{j=1}^{K}\omega(y_j)\,p^{-1}(y_i|x_j) \qquad i=1,2,\cdots,K \tag{4-21}$$

式中，$\omega(y_j)$由式（4-16）决定。

按式（4-21）算得的分布，若有某一 $q(x_i)$大于 1 或小于 0，说明前面计算的 $I(X;Y)$的极值不存在。这时最大值必然出现在边界上，可以令该 $q(x_i)=0$，用另外的方法再试算。

综上所述，计算信道容量 C 按下面的步骤进行。

（1）先验证信道转移概率矩阵 $\boldsymbol{P}=[p(y_j \mid x_i)]$是方阵，且矩阵 $\boldsymbol{P}$的行列式 $|[p(y_j \mid x_i)]| \neq 0$；

（2）计算出逆矩阵 $\boldsymbol{P}^{-1}=[p^{-1}(y_j \mid x_k)]$；

（3）根据式（4-17），计算出 $H(Y|x_i)=-\sum_{j=1}^{K}p(y_j|x_i)\ln p(y_j|x_i),\ i=1,2,\cdots,K$；

（4）根据式（4-18），计算出信道容量 C；

（5）验证是否满足 $q(x_i)\geqslant 0$，$i=1,2,\cdots,K$。

- 先由式（4-16）计算出$\omega(y_k)$，$k=1,2,\cdots,K$；
- 再由式（4-21）计算 $q(x_i),\ i=1,2,\cdots,K$。

【例 4.9】 给出信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{bmatrix}$，$K=2$，求信道容量 C。

（1）$\boldsymbol{P}$矩阵的行列式 $|\boldsymbol{P}|=\begin{vmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{vmatrix}\neq 0$，说明 $\boldsymbol{P}$是一个非奇异方阵，可按上述方法求信道容量；

（2）$\boldsymbol{P}$的逆矩阵 $\boldsymbol{P}^{-1}=\left[p^{-1}(y_j|x_i)\right]=\begin{bmatrix}8/7 & -1/7\\ -2/7 & 9/7\end{bmatrix}=\begin{bmatrix}1.143 & -0.143\\ -0.286 & 1.286\end{bmatrix}$；

（3）算出
$$\begin{cases} H(Y|x_1)=-\sum_{j=1}^{2}p(y_j|x_1)\ln p(y_j|x_1)=-0.9\ln 0.9-0.1\ln 0.1=0.325 \\ H(Y|x_2)=-\sum_{j=1}^{2}p(y_j|x_2)\ln p(y_j|x_{21})=-0.2\ln 0.2-0.8\ln 0.8=0.5 \end{cases};$$

（4）信道容量

$$\begin{aligned} C&=\ln\sum_{k=1}^{2}\exp\left\{-\sum_{i=1}^{2}p^{-1}(y_i|x_k)\,H(Y|x_i)\right\} \\ &=\ln[\exp(-1.143\times 0.325+0.143\times 0.5)+\exp(0.286\times 0.325-1.286\times 0.5)] \\ &=\ln[\mathrm{e}^{-0.3}+\mathrm{e}^{-0.55})=0.276 \text{（奈特/码符号）} \end{aligned}$$

（5）下面验证 $q(x_i)\geqslant 0$，$i=1,2$ 是否成立。

- 先根据 $\omega(y_k)=\exp\left\{-C-\sum_{i=1}^{2}p^{-1}(y_i|x_k)\,H(Y|x_i)\right\}$ $(k=1,2)$ 算出

$$\begin{cases}\omega_1 = \mathrm{e}^{-0.276-0.3} = 0.562\\ \omega_2 = \mathrm{e}^{-0.276-0.55} = 0.438\end{cases}$$

- 再算得 $\begin{cases} q(x_1) = \sum_{j=1}^{2} \omega(y_j) p^{-1}(y_1 \mid x_j) = 0.562 \times 1.143 - 0.438 \times 0.286 = 0.517 > 0 \\ q(x_2) = \sum_{j=1}^{2} \omega(y_j) p^{-1}(y_2 \mid x_j) = -0.562 \times 0.143 + 0.438 \times 1.286 = 0.483 > 0 \end{cases}$

说明上面算得的 $C = 0.276$（奈特/码符号）是正确的。

4.3 组合信道的容量

考虑有两个信道，如图 4-9 所示。

信道 1 $\begin{cases} \text{输入符号集：} X \in \{x_1, x_2, \cdots, x_i, \cdots\} \\ \text{输出符号集：} Y \in \{y_1, y_2, \cdots, y_j, \cdots\} \\ \text{信道转移概率矩阵：} \boldsymbol{P}_1 = [p(y_j \mid x_i)] \end{cases}$

信道 2 $\begin{cases} \text{输入符号集：} X' \in \{x_1', x_2', \cdots, x_i', \cdots\} \\ \text{输出符号集：} Y' \in \{y_1', y_2', \cdots, y_j', \cdots\} \\ \text{信道转移概率矩阵：} \boldsymbol{P}_2 = [p(y_j' \mid x_i')] \end{cases}$

下面介绍信道在三种不同组合情况下的信道容量。

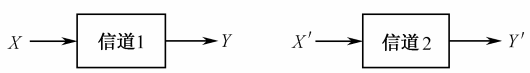


图 4-9　两个信道

4.3.1 独立并行信道

在这种情况下，两个信道作为一个信道使用，传送符号 XX'，接收符号 YY'，因为两个信道是独立的，故并行信道的转移概率为 $p\left(y_j y_j' \mid x_i x_i'\right) = p\left(y_j \mid x_i\right) p\left(y_j' \mid x_i'\right)$，这就相当于信道离散无记忆时应满足的条件，根据定理 2.4，对于离散无记忆信道，下式成立

$$I(XX';YY') \leqslant I(X;Y) + I(X';Y') \tag{4-22}$$

对上面的不等式两边取最大值，得

$$C \leqslant C_1 + C_2 \tag{4-23}$$

式（4-23）说明在两信道并行使用的情况下，总信道容量小于两信道单独使用时的信道容量之和。

式（4-23）中等号成立的条件即式（4-22）等号成立的条件，要求信源离散无记忆，满足 $q(x_i x_i') = q(x_i) q(x_i')$，即要求两信道独立使用且输入独立。

上面的结论可以推广到 N 个信道的并行组合，当 N 个信道并行独立使用时，记 C_k（$k = 1, 2, \cdots, N$）为第 k 个信道的信道容量，C 为组合信道的总容量，则有

$$C \leqslant \sum_{k=1}^{N} C_k \tag{4-24}$$

当 N 个信道独立输入且独立使用时，有

$$C=\sum_{k=1}^{N}C_k \tag{4-25}$$

4.3.2 和信道

两个信道轮流使用（不能同时），使用概率分别为 p_1, p_2，且 $p_1+p_2=1$，记概率分布 $\boldsymbol{P}=(p_1, p_2)$，和信道的平均互信息计算如下。

$$\begin{aligned}
&I(\boldsymbol{P})=I(p_1, p_2)\\
&=\sum_{x_i}\sum_{y_j}p_1q(x_i)p\left(y_j|x_i\right)\log\frac{p\left(y_j|x_i\right)}{p_1\omega(y_j)}+\sum_{x_i'}\sum_{y_j'}p_2q(x_i')p\left(y_j'|x_i'\right)\log\frac{p(y_j'|x_i')}{p_2\omega(y_j')}\\
&=\sum_{x_i}\sum_{y_j}p_1q(x_i)p\left(y_j|x_i\right)\log\frac{p\left(y_j|x_i\right)}{\omega(y_j)}+\sum_{x_i'}\sum_{y_j'}p_2q(x_i')p\left(y_j'|x_i'\right)\log\frac{p\left(y_j'|x_i'\right)}{\omega(y_j')}-p_1\log p_1-p_2\log p_2\\
&=p_1I(X;Y)+p_2I(X';Y')+H_2(\boldsymbol{P})
\end{aligned}$$

式中，$H_2(\boldsymbol{P})=-p_1\log p_1-p_2\log p_2$。

根据定义，有

$$\begin{aligned}
C&=\max_{\boldsymbol{P},q(x),q(x')}I\left(\boldsymbol{P}\right)\\
&=\max_{\boldsymbol{P},q(x),q(x')}\left[p_1I\left(X;Y\right)+p_2I\left(X';Y'\right)+H_2\left(\boldsymbol{P}\right)\right]\\
&=\max_{\boldsymbol{P}}\left[p_1C_1+p_2C_2+H_2(\boldsymbol{P})\right]
\end{aligned}\tag{4-26}$$

下面求使式（4-26）取极大值的 $\boldsymbol{P}$。

令 $\dfrac{\partial\left[p_1C_1+p_2C_2+H_2\left(\boldsymbol{P}\right)\right]}{\partial p_1}=0$，对数以 2 为底，注意到 $p_2=1-p_1$，得

$$C_1-C_2-\log p_1-\frac{1}{\ln 2}+\log p_2+\frac{1}{\ln 2}=0$$

记

$$C_1-\log p_1=C_2-\log p_2=\lambda\text{（}\lambda\text{为待定常数）} \tag{4-27}$$

从式（4-27）中解出

$$\begin{cases}p_1=2^{C_1-\lambda}\\ p_2=2^{C_2-\lambda}\end{cases} \tag{4-28}$$

将式（4-28）代入条件 $p_1+p_2=1$，得

$$\lambda=\log\left[2^{C_1}+2^{C_2}\right] \tag{4-29}$$

式（4-28）中的 p_1, p_2 就是使平均互信息量 $I(p_1, p_2)$达到最大的取值，将其代入式（4-26），得

$$\begin{aligned}
C&=2^{C_1-\lambda}C_1+2^{C_2-\lambda}C_2-2^{C_1-\lambda}\log 2^{C_1-\lambda}-2^{C_2-\lambda}\log 2^{C_2-\lambda}\\
&=2^{C_1-\lambda}\left[C_1-\left(C_1-\lambda\right)\right]+2^{C_2-\lambda}\left[C_2-\left(C_2-\lambda\right)\right]\\
&=\lambda(2^{C_1-\lambda}+2^{C_2-\lambda})\\
&=\lambda\text{（}p_1+p_2\text{）}=\lambda
\end{aligned}\tag{4-30}$$

将式（4-29）代入式（4-30），得

$$C = \log\left[2^{C_1} + 2^{C_2}\right]$$

上述结果也可推广到 N 个信道轮流使用的情况，当 N 个信道以不同概率轮流使用时，记 C_k（$k = 1, 2, \cdots, N$）为第 k 个信道的信道容量，C 为组合信道的总容量，则有

$$C = \log \sum_{k=1}^{N} 2^{C_k} \tag{4-31}$$

每个信道的使用概率为 $p_k = 2^{C_k - C}, k = 1, 2, \cdots, N$。

4.3.3 串行信道

将两个信道级联，有 $X' = Y$，如图 4-10 所示，要求信道 1 的输出等于信道 2 的输入。

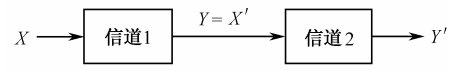


图 4-10　串行信道

串行信道的信道转移概率 $p(y'_j | x_i) = \sum_j p(y_j | x_i) p(y'_j | x'_i = y_j)$

用矩阵表示为

$$\left[p(y'_j | x_i)\right] = \left[p(y_j | x_i)\right] \cdot \left[p(y'_j | x'_i)\right] \tag{4-32}$$

式（4-32）中，$\left[p(y'_j | x_i)\right]$ 是串联信道的总信道转移概率矩阵，$\left[p(y_j | x_i)\right]$ 和 $\left[p(y'_j | x'_i)\right]$ 分别是第一个和第二个信道的信道转移概率矩阵。

【例 4.10】　给定两个信道，信道转移概率矩阵分别为

$$\boldsymbol{P}_1 = \left[p\left(y_j | x_i\right)\right] = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{bmatrix} \qquad \boldsymbol{P}_2 = \left[p\left(y'_j | x'_i\right)\right] = \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix}$$

采用图 4-10 所示的方式级联使用，根据式（4-32）可求得串行信道的信道转移概率矩阵为

$$\boldsymbol{P} = \boldsymbol{P}_1 \cdot \boldsymbol{P}_2 = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.5 & 0.2 & 0.3 \end{bmatrix} \cdot \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.49 & 0.51 \\ 0.41 & 0.59 \end{bmatrix}$$

可见，串行级联信道的信道转移概率趋向于两个独立信道转移概率的均值，这是很不利的情况，这种情况下出错概率增大，传信能力减小。

求得了串联信道的总信道转移概率矩阵，即可利用前面介绍的方法求总信道容量。

若将 N 个转移概率相同的信道级联，可证明当 $N \to \infty$ 时，其总信道容量将趋于零，参见习题 4.18。

对于例 4.10 给出的结论，可用数据处理定理说明。信道 1 的统计特性用 $\boldsymbol{P}_1 = [p\ (y \mid x)]$ 表示，信道 2 的统计特性用 $\boldsymbol{P}_2 = [p\ (y' \mid x')]$ 表示，因为信道 1 和信道 2 是独立的，信道 2 的输出 Z 只与其输入 Y 及信道转移概率 $\boldsymbol{P}_2 = [p\ (y' \mid x')]$ 有关，而与 X 无关。因此信道 1 和信道 2 串连就构成了一个马尔可夫链，对于马尔可夫链有如下定理。

定理 4.4　若随机变量 X, Y, Z 组成一个马尔可夫链，如图 4-11 所示，则有

$$I(X;\ Z) \leqslant I(X;\ Y) \tag{4-33}$$

$$I(X;\ Z) \leqslant I(Y;\ Z) \tag{4-34}$$

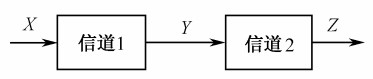


图 4-11　马尔可夫链

证明式（4-33）：

$$\begin{aligned}
&I(X;\ Z) - I(X;\ Y)\\
&= [H(X) - H(X \mid Z)] - [H(X) - H(X \mid Y)]\\
&= H(X \mid Y) - H(X \mid Z)\\
&\leqslant H(X \mid Y) - H(X \mid YZ) \qquad \text{（条件熵小于等于无条件熵）}\\
&= H(X \mid Y) - H(X \mid Y) \qquad \text{（因为 } X, Y, Z \text{ 为马尔可夫链）}\\
&= 0
\end{aligned}$$

则　$I(X;\ Z) \leqslant I(X;\ \mathrm{Y})$　证毕

式（4-33）说明，信道 2 对于从 Z 获得关于 X 的信息毫无帮助，通过信道后不会使信息增加。若将信道 2 看成一个数据处理系统，例如通信系统中常用到的采样、量化、编码及译码等装置，那么上述结论称为**数据处理定理**，即无论经过何种数据处理，都不会使信息量增加。

式（4-34）可类似证得。

从上面的证明过程可看出，满足 $H(X \mid Z) = H(X \mid Y)$，即满足

$$p\ (x \mid z) = p\ (x \mid y) \tag{4-35}$$

时，式（4-33）中的等号成立，即有 $I\ (X; Z) = I\ (X; \mathrm{Y})$，说明这种情况下串联传输不会增加信息的损失。

【例 4.11】　两个离散信道 $\boldsymbol{P}_1 = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}$，$\boldsymbol{P}_2 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$，将它们串行连接使用，如图 4-10 所示，计算总信道容量 C。

（1）先计算总信道的信道转移概率矩阵

$$\boldsymbol{P} = \boldsymbol{P}_1 \cdot \boldsymbol{P}_2 = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}$$

可见该串联信道的总信道矩阵 $\boldsymbol{P}$ 等于第一级信道的信道矩阵 $\boldsymbol{P}_1$，从而概率分布满足

$$p\ (y \mid x)\ =\ p\ (z \mid x)\text{（对所有的 } x, y, z\text{）} \tag{4-36}$$

对式（4-36）两边关于 x 求和，得

$$\sum_x q(x)\, p(y \mid x) = \sum_x q(x)\, p(z \mid x)$$

$$p\ (y) = p\ (z) \tag{4-37}$$

利用式（4-37），由式（4-36）得

$$\frac{p(y|x)p(x)}{p(y)}=\frac{p(z|x)p(x)}{p(z)}$$

$$p(x\mid y)=p(x\mid z)$$

上式恰好就是式（4-35）所要求的条件，这就说明，不论信源符号 $X=\{x\}$ 如何分布，只要串联信道的总信道转移概率矩阵等于第一个信道的转移概率矩阵，这样的串联噪声信道就不会增加信道的信息损失，总的信道容量就等于第一个信道的信道容量。

（2）计算信道容量 C

在例 4.11 中，第一个信道是输入只有两个消息的情况，设最佳分布为 $q(x_1)=\alpha$，$q(x_2)=1-\alpha$，仿照例 4.8 可算出 $\alpha=0.4$，则信道容量 $C=C_1=0.32$（比特/符号）。

本 章 小 结

本章主要定义了信道容量，讨论了信道容量的计算方法，信道容量是信息在有噪信道中传输时，无失真条件下信道所能容纳的信息率的极限值。信道容量是信息论中最富有贡献的概念之一，研究信道，其核心问题就是研究信道容量及信源最佳分布。

信道容量定义为平均互信息量在信源最佳分布下取得的最大值，即

$$C\triangleq\max_{q(x)}I(X;Y)$$

文中讨论并证明了使平均互信息量达到信道容量的充要条件，并给出如下几种情况下信道容量的计算方法。

（1）准对称信道；

（2）信源只含两个消息；

（3）信道转移概率矩阵为可逆方阵。

至于一般情况下，信道容量的求解是件冗长烦琐的事情，可采用叠代算法利用计算机编程计算，这在不少书中都有介绍，读者可参考其他书籍。

本章还讨论了多个信道组合使用情况下，总信道容量的计算方法，讨论了以下几种情况。

（1）N 个信道独立并行使用：记每个信道单独使用时的信道容量为 C_k，k=1, 2, … , N，则总信道容量 C 满足 $C\leq\sum_{k=1}^{N}C_k$，当 N 个信道独立输入且独立使用时等号成立。

（2）N 个信道轮流使用：各信道使用概率为 p_k，各信道的容量为 C_k，k =1,2, …, N，总信道容量为 $C=\log\sum_{k=1}^{N}2^{C_k}$，每个信道的使用概率为 $p_k=2^{C_k-C},k=1,2,\cdots,N$。

（3）N 个信道串联使用：记各个信道的信道转移概率矩阵为 $\boldsymbol{P}_k$，k =1,2, …, N，则总信道的信道转移概率矩阵 $\boldsymbol{P}$ 等于各信道转移概率矩阵相乘，即 $\boldsymbol{P}=\boldsymbol{P}_1\boldsymbol{P}_2\cdots\boldsymbol{P}_N$，矩阵的乘法要满足：左乘矩阵的列数应等于右乘矩阵的行数，且矩阵相乘不满足交换率。计算出 $\boldsymbol{P}$ 后，再根据本章介绍的方法计算信道容量。

思考题与习题

4.1 设信道输入符号集 $X=\{x_1,x_2,\cdots,x_k\}$，输出符号集 $Y=\{y_1,y_2,\cdots,y_s\}$，如果信道是有噪无损信道，则其信道容量为多少？如果信道是无噪确定信道，则其信道容量又为多少？

4.2 信源的最佳编码使信道码符号等概分布，而且平均码长最短，这种说法对吗？

4.3 信道的信息传输率是信道输入分布的函数，但信道容量与信道的输入分布无关，这种说法对吗？

4.4 信道容量是关于信源分布（即信道输入分布）求最佳，但是信道容量与信源分布无关，它仅是信道转移概率的函数，这种说法对吗？

4.5 判断以下几种信道是不是准对称信道

（1）$\begin{bmatrix}0.2 & 0.3 & 0.5\\ 0.5 & 0.2 & 0.3\end{bmatrix}$　（2）$\begin{bmatrix}0.7 & 0.3\\ 0.4 & 0.6\\ 0.3 & 0.7\end{bmatrix}$　（3）$\begin{bmatrix}0.7 & 0.1 & 0.2\\ 0.2 & 0.1 & 0.7\end{bmatrix}$　（4）$\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}\end{bmatrix}$

4.6 证明无损信道的充要条件是信道转移概率矩阵中每一列有且只有一个非零元素。

4.7 计算以下离散无记忆信道 DMC 的容量及最佳分布。

（1）$\boldsymbol{P}=\begin{bmatrix}1-p & p & 0\\ 0 & 1-p & p\\ p & 0 & 1-p\end{bmatrix}$　（2）$\boldsymbol{P}=\begin{bmatrix}\frac{1-p}{2} & \frac{1-p}{2} & \frac{p}{2} & \frac{p}{2}\\ \frac{p}{2} & \frac{p}{2} & \frac{1-p}{2} & \frac{1-p}{2}\end{bmatrix}$

（3）$\boldsymbol{P}=\begin{bmatrix}1/2 & 1/2 & 0 & 0 & 0\\ 0 & 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/2 & 1/2 & 0\\ 0 & 0 & 0 & 1/2 & 1/2\\ 1/2 & 0 & 0 & 0 & 1/2\end{bmatrix}$　（4）$\boldsymbol{P}=\begin{bmatrix}1/2 & 1/2 & 0\\ 1/4 & 1/4 & 1/2\end{bmatrix}$

（5）$\boldsymbol{P}=\begin{bmatrix}1/3 & 2/3\\ 2/3 & 1/3\end{bmatrix}$　（6）$\boldsymbol{P}=\begin{bmatrix}1/3 & 1/3 & 1/6 & 1/6\\ 1/6 & 1/3 & 1/3 & 1/6\end{bmatrix}$

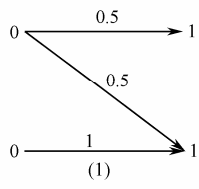
4.8 计算题图 4-1 所示离散无记忆信道 DMC 的容量及最佳分布。

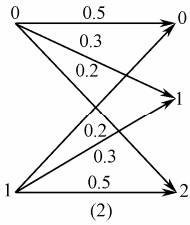
4.9 对题图 4-2 所示二进制删除信道，有 $q(x=0)=\alpha$，$q(x=1)=1-\alpha$，求

（1）平均互信息量 $I(X;\ Y)$;

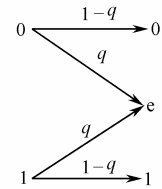
（2）α为何值时，$I(X;Y)$ 达到最大值 C;

（3）根据（2）中的α 值，计算 $I(x;\ y)$ 值 $I(0;0), I(1;0), I(0;\mathrm{e})$。





题图 4-1　离散无记忆信道



题图 4-2　二进制删除信道

4.10　给定离散信道的信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-p & p & 0 & 0\\ p & 1-p & 0 & 0\\ 0 & 0 & 1-q & q\\ 0 & 0 & q & 1-q\end{bmatrix}$，计算其信道容量 C。

4.11　给定离散信道 $\boldsymbol{P}=\begin{bmatrix}0.3 & 0.7\\ 0.5 & 0.5\end{bmatrix}$，计算信道容量 C。

4.12　离散无记忆信道，其信道转移概率矩阵 $\left[p(y|x)\right]=\begin{bmatrix}1-\delta-\varepsilon & \varepsilon & \delta\\ \delta & \varepsilon & 1-\delta-\varepsilon\end{bmatrix}$，求

（1）信道容量 C 及最佳分布 $q(x)$；

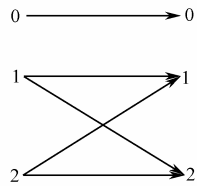
（2）最佳分布时的 $I(x;y)$=$I(0;0)$；$I(1;0)$。

4.13　信道传输 0 和 1 两个符号，$p(0)=p(1)=0.5$，传输错误概率 $p_e=0.01$，信息传输率 $R_t=1000$ b / s，求此信道的信道容量。

4.14　求题图 4-3 所示离散信道的信道容量 C，及对应的输入最佳分布。并求出 $\varepsilon=0$，$\varepsilon=1$ 和 $\varepsilon=0.5$ 时的信道容量。

4.15　计算例 4.11 中串联信道的总信道容量 C。

4.16　将两个信道转移概率矩阵都为 $\boldsymbol{P}=\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$ 的信道串联使用，第一个信道的输入符号为 $X=\{x_0, x_1, x_2, x_3\}$，等概分布，求平均互信息量 $I(X;Z)$ 和 $I(X;Y)$，并对二者进行比较。



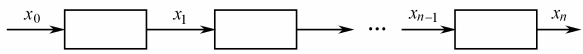
题图 4-3　离散信道

4.17　一个离散无记忆模 K 加信道：输入、输出字符集都为{0，1，…，K−1}，输出 y 等于输入 x 模 K 加上噪声 z，z 为分布在{0，1，…，K−1}上的随机变量。

（1）证明：$I(X;Y)=H(Y)-H(Z)$;

（2）求以 $H(Z)$为函数的信道容量及最佳输入分布。

4.18　n 个同样的二进制对称信道 BSC 级联，如题图 4-4 所示，各信道的转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}p & 1-p\\ 1-p & p\end{bmatrix}$，证明它等价于一个转移概率为 $\frac{1}{2}\left[1-(1-2p)^n\right]$ 的 BSC，且当 $n\to\infty$ 时，信道容量 $C\to 0$。



题图 4-4　n 个二进制信道级连

4.19　二元对称信道，固有误码率为 $p=0.01$，设该信道以 R_t=1000 b/s 的速率传输二进制输入符号，输入符号概率分布 $p(0)=p(1)=0.5$，求此信道的信道容量。

4.20　二元对称信道，固有误码率为 $p=0.02$，设该信道以 R_t=1500 b/s 的速率传输输入符号，一消息序列共有 14 000 个二元符号，其中 $p(0)=p(1)=0.5$，试问 10 s 内是否能将该消息序列无失真地传输完？

第 5 章
有噪信道编码

内容提要

本章介绍了信道编码和译码的基本概念，介绍了两种常用的译码准则：最大后验概率译码准则和极大似然译码准则，还介绍了在这两种译码准则下错误概率的计算方法。

本章以较大的篇幅介绍了信道编码定理及信道编码逆定理。Shannon 1948 年提出并证明了信道编码的正、反定理，揭示了信道的输入在信道传输率小于信道容量（$R<C$）的条件下，可以以任意小的错误概率抵达信宿。同时又指出在 $R>C$ 的条件下，则无论如何也不可能实现无差错传输。后来许多研究者给出了许多更严格的证明，并给出了各种信道及编码条件下的错误概率的上、下限，这些都为设计合理的通信系统提供了理论依据。

在证明信道编码逆定理的过程中，用到了费诺（Fnao）不等式，Fnao 不等式本身就是信息论中的一个重要不等式。

知识要点

译码规则，错误概率，信道编码定理，信道编码逆定理，费诺不等式。

教学建议

对本章介绍的两种译码准则：最大后验概率译码准则和极大似然译码准则，讲述在什么前提及在什么场合应用才是最佳译码准则。鉴于信道编码定理及其逆定理的重要性，文中给予了证明，若课时不够可略去不讲，但应要求学生弄懂定理的物理意义。建议学时数为 5 学时。

5.1 信道编码的基本概念

在有噪信道中传输信息，我们总是希望信息传输快捷可靠，但由于信道的信息传输率受到信道容量的限制，不可能无穷大，另一方面，由于信息在信道中传输时不可避免地会拾取各种噪声干扰，所以传输误差不可能为零。我们要做的是通过各种编码方法尽可能地提高信息传输率，并将传输误差控制在可接受的范围内。

在第 3 章中讨论了对信源的编码，信源编码以提高传输效率为主要考虑因素，编码时应尽量压缩信源冗余度。信道编码以提高传输可靠性为主要考虑因素，本章讨论信道编码的一些基本概念及信道编码定理。

信道有多种形式，如电视、广播、微波、电话、计算机局域网和宽带网等，不失一般性，可将信道用图 5-1 所示的模型表示。

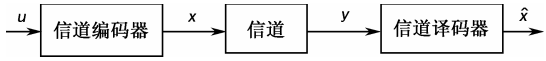


图 5-1　信道模型

衡量信道传输快慢的指标是信息传输率，而衡量信息传输可靠性的指标是平均误码率，误码率与信道的统计特性（用信道转移概率表示）有关，改变信道的统计特性成本太高，所以可事先对信源编码器输出的符号序列按照某种规则进行编码，一般的方法是给信源序列加上一定的冗余度，这种编码谓之信道编码，编好的代码称为码字，将码字送入信道传输。而在信道输出端，信道译码器根据编码规则对信道输出符号进行估值，尽量使这种估值接近输入码字。

从上面的叙述可看出，选择合适的编码规则是信道编码的关键，这部分内容将在 5.2 节中讨论。

上面的过程可根据图 5-1 表述如下。

信源输出序列 $\boldsymbol{u}$，经信道编码器编成码字 $\boldsymbol{x}=f(\boldsymbol{u})$ 并输入信道，由于干扰，信道输出 $\boldsymbol{y}$，信道译码器对 $\boldsymbol{y}$ 估值得 $\hat{\boldsymbol{x}}=F(\boldsymbol{y})$，译码规则 F 是对编码规则 f 的一种逆变换。

作为一般性介绍，下面举例介绍几种简便的编码方法。

【例 5.1】　给定二进制对称信道，如图 1-5 所示，信道固有误码率为 p（p<0.5）。

编码规则：为提高可靠性，每个信道符号重复三次发送。

译码规则：择多译码，即信宿方收到的三个符号中有两个或三个为 1，就将此次接收符号判决为 1；若三个符号中有两个或三个为 0，就将此次接收符号判决为 0。

图 5-2 所示为重复编码传输示意图，计算错误概率 p_e。

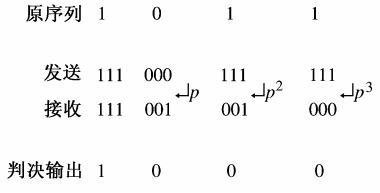


图 5-2　重复编码传输示意图

信源输出序列为 $\boldsymbol{u}$={1，0，1，1}，将信源符号重复三次发送，即向信道输入序列 $\boldsymbol{x}$={111，000，111，111}，这一过程称为信道编码。在信道传输过程中，由于 p 的存在，传输出错，信道输出为 $\boldsymbol{y}$ ={111，001，001，000}。

第一种情况：111→111　无错

第二种情况：000→001　错 1 位

第三种情况：111→001　错 2 位

第四种情况：111→000　错 3 位

根据译码规则，估值输出 $\hat{\boldsymbol{y}}$ = {1，0，0，0}，这一过程称为信道译码。与原发送序列相比较，可见，如三位里面错两位或三位里面错三位，则属于不可纠错误。

下面计算错误概率：假设信道离散无记忆，即 $p(\boldsymbol{y}|\boldsymbol{x})=\prod_{i=1}^{3}p(y_i|x_i)$，错误概率 p_e 可如下计算：

$$p_e=\binom{3}{2}p^2(1-p)+\binom{3}{3}p^3=3p^2-2p^3<p \tag{5-1}$$

式中，$\binom{n}{m}=\frac{m!}{n!(m-n)!}$ 表示组合，即 m 个符号中每次取 n 个。

可见，重复编码的结果使错误概率下降。

如果提高重复次数，能否进一步使错误概率下降呢，见下面的例子。

【例 5.2】　离散无记忆二进制对称信道，信道固有误码率为 p（p<0.5），为提高可靠性，每个信道符号重复 $2n+1$ 次发送，择多译码，即在 $2n+1$ 个接收码符号中，若有 $n+1$ 个以上 1 就将此次接收估值为 1，若有 $n+1$ 个以上 0 就将此次接收估值为 0，计算错误概率 p_e。

显然，在 $2n+1$ 个接收码符号中，错 $n+1$ 个及其以上属于不可纠错码。

错误概率

$$\begin{aligned}p_e&=\sum_{k=n+1}^{2n+1}\binom{2n+1}{k}p^k(1-p)^{2n+1-k}\\&<n\binom{2n+1}{n+1}p^{n+1}(1-p)^{2n+1-(n+1)}\\&=n\binom{2n+1}{n+1}p^{n+1}(1-p)^n\xrightarrow{n\to\infty}0\end{aligned}$$

从上述表达式可看出，随着 $n\to\infty$，将有 $p_e\to 0$。

可以根据式 $p_e=n\binom{2n+1}{n+1}p^{n+1}(1-p)^n$，由给定的错误概率 p_e，选择重复发送次数 n。

【例 5.3】　逆重复码

离散无记忆二进制对称信道，固有误码率为 p（p<0.5），信源输出序列为三位二进制数字。为提高传输效率，将信源输出的三位二进制序列经编码后，仅向信道发送一位。

编码规则：预先将信源输出序列进行择多编码，即若信源输出的三位符号中有两位或三位是 1，就将此信源序列编码为 1，若三位符号中有两位或三位是 0，就将此信源序列编码为 0。

译码规则：将接收到的一位符号重复三次译出，即若接收到 1 就译码为 111，若接收到 0 就译码为 000。

图 5-3 所示为逆重复编码传输示意图，计算错误概率 p_e。分两步进行：

图 5-3　逆重复编码传输示意图

（1）先设 $p=0$，计算这种编码方法带来的固有误比特率 p_1。

信道输入符号集 $X=\{000,001,010,011,100,101,110,111\}$；

判决输出符号集 $Y=\{000,111\}$。

译码规则 $\begin{cases} F(000,\ 001,\ 010,\ 100)=000 \\ F(011,\ 101,\ 110,\ 111)=111 \end{cases}$

因为后验概率 $\begin{cases} \phi(\boldsymbol{x}|\boldsymbol{y})=\phi(000\,|\,000)=\dfrac{1}{4} \\ \phi(\boldsymbol{x}|\boldsymbol{y})=\phi(111\,|\,111)=\dfrac{1}{4} \end{cases}$

则错误概率 $\begin{cases} p(e|y)=p(e|000)=1-\dfrac{1}{4}=\dfrac{3}{4} \\ p(e|y)=p(e|111)=1-\dfrac{1}{4}=\dfrac{3}{4} \end{cases}$

假设 8 组输入序列是等概发送的，由于信道的对称性，两个估值序列也是等概分布的，则每个序列的平均错误概率为 $0.5p(e|000)+0.5p(e|111)=\dfrac{3}{4}$，误比特率 $p_1=\dfrac{1}{3}\times\dfrac{3}{4}=\dfrac{1}{4}$。

（2）再设 $p\neq 0$，计算由于信道噪声引起的误比特率 p_2。

因为每个序列有三位二进制数字，但只发送一位，这一位的出错概率为 p，故序列差错概率为 p，误比特率 $p_2=\dfrac{1}{3}p$。

（3）总误比特率

$$p_e=p_1+p_2=\frac{1}{4}+\frac{1}{3}p$$

【例 5.4】　奇偶校验码

在信息序列后面加上一位校验位，使之模 2 和等于 1，这样的编码称为奇校验码；若使模 2 和等于 0，这样的编码就称为偶校验码，即每个码矢中 1 的个数固定为奇数或偶数。

这两种码都能检验出奇数位错误，但无法判断错的是码字中的哪一位，故没有纠错能力。这种编码方式可应用在有反馈信道的场合，一旦发现有错，便利用反馈信道要求重发。

在上面这 4 个例子中，例 5.3 是为了提高传输效率，例 5.1、例 5.2 和例 5.4 都是为了提高传输质量。可见，编码问题主要要考虑的是如何提高传输效率和传输质量，信道纠错编码以提高传输质量为主。

5.2 译码规则及错误概率

信道总不可避免地会掺杂噪声，所以信息在信道传输过程中，差错是不可避免的。错误概率与信道统计特性有关，选择合适的译码规则可以弥补信道的不足。

在例 2.17 中提到过，给定信道 $\boldsymbol{P}=\begin{bmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{bmatrix}$，当 $\varepsilon=1$ 时，信道转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$，对应的信道如图 5-4 所示，这是强噪信道，如果按常规，收到 0 判决为 0，收到 1 判决为 1，显然完全错了。如果译码规则定为收到 0 判决为 1，收到 1 判决为 0，则反而正确。可见译码规则对传输系统错误概率的影响是很大的。

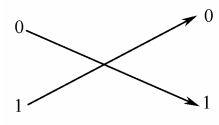


图 5-4 强噪信道

下面介绍两种典型的译码规则。

1．最大后验概率译码准则

设信源共有 M 个消息，假设信源编码器已用 M 个码矢 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m, \cdots, \boldsymbol{x}_M$ 对它进行了最佳编码。

编码后发送码矢 $\boldsymbol{x}_k$，其发送概率为 $q(\boldsymbol{x}_k)$，通过信道转移概率为 $p(\boldsymbol{y}|\,\boldsymbol{x}_k)$的信道传输，接收到矢量 $\boldsymbol{y}$，信道译码器输出 $\hat{\boldsymbol{x}}_k$，$\hat{\boldsymbol{x}}_k$ 是信道译码器根据事先定出的译码规则，对接收矢量 $\boldsymbol{y}$ 的一个估值，通信过程可用图 5-5 所示框图表示。

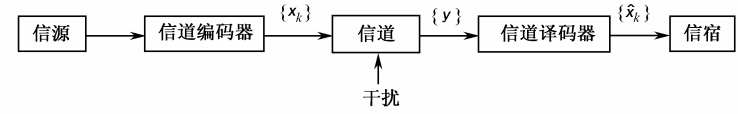


图 5-5 通信过程框图

显然，当估值 $\hat{\boldsymbol{x}}_k \neq \boldsymbol{x}_k$ 时，就产生了误码，用$\phi(\boldsymbol{x}|\boldsymbol{y})$表示后验概率，则收到 $\boldsymbol{y}$ 估错的概率为

$$p_{\mathrm{e}}(\boldsymbol{y})=\phi\left(\hat{\boldsymbol{x}}_k \neq \boldsymbol{x}_k\,|\,\boldsymbol{y}\right)=\sum_{i\neq k}\phi\left(\boldsymbol{x}_i\,|\,\boldsymbol{y}\right)=1-\phi\left(\boldsymbol{x}_k\,|\,\boldsymbol{y}\right) \tag{5-2}$$

通信总希望错误概率最小，由式（5-2）可看出错误概率 $p_{\mathrm{e}}(\boldsymbol{x}_k)$最小等同于后验概率 $\phi(\boldsymbol{x}_k|\,\boldsymbol{y})$最大，这就是**最大后验概率译码准则**。

根据概率关系式

$$\phi(\boldsymbol{x}\,|\,\boldsymbol{y})=\frac{p(\boldsymbol{x}\boldsymbol{y})}{\omega(\boldsymbol{y})} \tag{5-3}$$

由于最大后验概率译码准则的意思是收到矢量 $\boldsymbol{y}$ 后，在所有的 $\boldsymbol{x}_m$（$m=1, 2, \cdots, M$）中，选一个后验概率$\phi(\boldsymbol{x}_m|\boldsymbol{y})$最大的 $\boldsymbol{x}_m$ 值，作为对 $\boldsymbol{y}$ 的估值 $\hat{\boldsymbol{x}}_k=\boldsymbol{x}_k$，那么对这 M 个 $\boldsymbol{x}_m$ 值来说，概率$\omega(\boldsymbol{y})$是同一个值，所以根据式（5-3），后验概率$\phi(\boldsymbol{x}|\boldsymbol{y})$最大就意味着全概率 $p(\boldsymbol{x}\,\boldsymbol{y})$最大，因此最大后验概率译码准则也称为**最大联合概率译码准则**。

【例 5.5】　信源分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.1 & 0.4 \end{bmatrix}$，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$，信道输出符号 $Y=\{y_1, y_2, y_3\}$，按最大后验概率准则译码。

（1）根据 $p(xy)=p(y|x)\,q(x)$ 算出全概率，用矩阵表示 $\left[p(xy)\right]=\begin{bmatrix} \underline{0.25} & \underline{0.15} & 0.10 \\ 0.01 & 0.07 & 0.02 \\ 0.12 & 0.12 & \underline{0.16} \end{bmatrix}$；

（2）根据 $\omega(y_j)=\sum_i p(x_i y_j)$，算出 $[\omega(y)]=[0.38 \quad 0.34 \quad 0.28]$；

（3）再由 $\phi(x|y)=\dfrac{p(xy)}{\omega(y)}$ 算出后验概率，用矩阵表示

$$\left[\phi(x|y)\right]=\begin{bmatrix} \underline{0.25/0.38} & \underline{0.15/0.34} & 0.10/0.28 \\ 0.01/0.38 & 0.07/0.34 & 0.02/0.28 \\ 0.12/0.38 & 0.12/0.34 & \underline{0.16/0.28} \end{bmatrix}$$

（4）按最大后验概率准则译码，在后验概率矩阵中，每列选一最大值（矩阵中带下划线的值），译为 $\begin{cases} y_1 \to x_1 \\ y_2 \to x_1 \\ y_3 \to x_3 \end{cases}$；

（5）若按最大联合概率译码准则译码，则在全概率矩阵[$p(xy)$]中每列选一最大值（矩阵中带下划线的值），也可译出 $\begin{cases} y_1 \to x_1 \\ y_2 \to x_1 \\ y_3 \to x_3 \end{cases}$。

可见这两种方法得到同一结果，因为要得到后验概率，计算步骤更多，所以可直接应用最大联合概率译码准则译码。

2．极大似然译码准则

前面介绍的最大后验概率译码准则等同于最小传输错误概率准则，所以从错误概率最小这一角度来说，这一译码准则是最好的。但在实际应用中，经常要用同一信道来传输各种信源的消息，一般总是知道信道的统计特性 $p(\boldsymbol{y}|\boldsymbol{x})$，而不知道信源分布 $q(\boldsymbol{x})$，因而全概率 $p(\boldsymbol{x}\,\boldsymbol{y})=p(\boldsymbol{y}|\boldsymbol{x})\,q(\boldsymbol{x})$无从求得。在这种情况下，我们可按最大信道转移概率来确定估值 $\hat{\boldsymbol{x}}$，即在收到矢量 $\boldsymbol{y}$ 后，在所有的 $\boldsymbol{x}_m$（$m=1, 2, \cdots, M$）中，选一个转移概率 $p(\boldsymbol{y}|\boldsymbol{x}_m)$最大的 $\boldsymbol{x}_m$ 值，作为对 $\boldsymbol{y}$ 的估值 $\hat{\boldsymbol{x}}_k=\boldsymbol{x}_k$，这一译码规则称为**极大似然译码规则**。

可以证明，在信道输入等概条件下，极大似然译码规则也是最佳的，因为若下式成立

$$p(\boldsymbol{y}|\boldsymbol{x}_k)>p(\boldsymbol{y}|\boldsymbol{x}_m) \qquad m=1, 2, \cdots, M$$

则信道输入等概时，有 $q(\boldsymbol{x}_k)=q(\boldsymbol{x}_m)$，代入上式得

$$p(\boldsymbol{y}|\boldsymbol{x}_k)\,q(\boldsymbol{x}_k)>p(\boldsymbol{y}|\boldsymbol{x}_m)\,q(\boldsymbol{x}_m)$$

有

$$p(\boldsymbol{x}_k\,\boldsymbol{y})>p(\boldsymbol{x}_m\,\boldsymbol{y})$$

说明全概率最大，对应最大联合概率译码准则，因此也是最佳的。

3．平均错误概率

式（5-2）是信道输出 $\boldsymbol{y}$ 而信道译码器估错的概率，对式（5-2）两边关于 $\boldsymbol{y}$ 求统计平均值，得

$$
\begin{aligned}
p_{\mathrm{e}} &= \sum_{j=1}^{M} \omega(\boldsymbol{y}_j)\, p_{\mathrm{e}}\left(\boldsymbol{y}_j\right) \\
&= \sum_{j=1}^{M} \omega(\boldsymbol{y}_j)\left[1-\phi\left(\boldsymbol{x}_k \mid \boldsymbol{y}\right)\right] \\
&= 1-\sum_{j=1}^{M} \omega(\boldsymbol{y}_j)\phi\left(\boldsymbol{x}_k \mid \boldsymbol{y}\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M} p(\boldsymbol{x}_i\boldsymbol{y}_j) - \sum_{j=1}^{M} p(\boldsymbol{x}_k\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x}}\sum_{\boldsymbol{y}} p(\boldsymbol{x}\boldsymbol{y}) - \sum_{\boldsymbol{y}} p(\boldsymbol{x}_k\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x}-\boldsymbol{x}_k}\sum_{\boldsymbol{y}} p(\boldsymbol{x}\boldsymbol{y}) \\
p_{\mathrm{e}} &= \sum_{\boldsymbol{x}-\boldsymbol{x}_k}\sum_{\boldsymbol{y}} p(\boldsymbol{x}\boldsymbol{y}) \qquad (5\text{-}4)
\end{aligned}
$$

式（5-4）表示平均误码率等于对全概率 $p(\boldsymbol{x}\,\boldsymbol{y})$ 求和，$\boldsymbol{y}$ 对所有取值求和，$\boldsymbol{x}$ 对 $\boldsymbol{X}=\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m, \cdots, \boldsymbol{x}_M\}$ 中除 $\boldsymbol{x}_k$ 以外的所有项求和，而 $\boldsymbol{x}_k$ 正是根据最大后验概率译码准则被选为 $\boldsymbol{y}$ 的估值 $\hat{\boldsymbol{x}}_k=\boldsymbol{x}_k$ 的那一项。

【例 5.6】 计算例 5.5 的平均错误概率，若信源等概分布，则对其译码，并求平均错误概率。

（1）求平均错误概率 p_{e}

根据式（5-4）得

$$p_{\mathrm{e}} = \sum_{\boldsymbol{x}-\boldsymbol{x}_k}\sum_{\boldsymbol{y}} p(\boldsymbol{x}\boldsymbol{y}) = 0.01+0.12+0.07+0.12+0.1+0.02 = 0.44$$

（2）当信源等概分布，按最大似然函数译码准则译码，例 5.5 已给出信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix} \underline{0.5} & 0.3 & 0.2 \\ 0.1 & \underline{0.7} & 0.2 \\ 0.3 & 0.3 & \underline{0.4} \end{bmatrix}$，在矩阵的每列中选一最大值（矩阵中带下划线的值），译码为

$$\begin{cases} y_1 \to x_1 \\ y_2 \to x_2 \\ y_3 \to x_3 \end{cases}$$

平均错误概率 $\qquad p_{\mathrm{e}} = \dfrac{1}{3}\left(0.1+0.3+0.3+0.3+0.2+0.2\right) = 0.467$

【例 5.7】 仍考虑例 5.1 给出的重复码，信源等概分布，采用极大似然译码规则进行译码，并计算平均错误概率 p_{e}。

根据编码规则 $\boldsymbol{x}=f(\boldsymbol{u})$，重复 3 次编码，即 $\begin{cases} \boldsymbol{x}_1 = f(0) = 000 \\ \boldsymbol{x}_2 = f(1) = 111 \end{cases}$。

原信道是固有误码率为 p（p<0.5）的二进制对称信道，其信道转移概率矩阵为

$\boldsymbol{P}=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$，经编码后得到的码字是码长为 3 的码矢 $\boldsymbol{x}=\{000，111\}$，送入 3 次扩展信道传输，因信道离散无记忆，故扩展信道的信道转移概率矩阵为

$$\boldsymbol{P}^{(3)}=\begin{array}{c}\\ 000\\ \\ 111\end{array}\begin{array}{c}\begin{array}{cccccccc}000 & 001 & 010 & 011 & 100 & 101 & 110 & 111\end{array}\\ \begin{bmatrix}\underline{(1-p)^3} & \underline{(1-p)^2p} & \underline{(1-p)^2p} & (1-p)p^2 & \underline{(1-p)^2p} & (1-p)p^2 & (1-p)p^2 & p^3\\ p^3 & (1-p)p^2 & (1-p)p^2 & \underline{(1-p)^2p} & (1-p)p^2 & \underline{(1-p)^2p} & \underline{(1-p)^2p} & \underline{(1-p)^3}\end{bmatrix}\end{array}$$

按最大后验概率译码准则译码，注意到 $p<0.5$，在矩阵的每一列中选一最大值（矩阵中带下划线的值），译码为

F（000）= 000，　F（001）= 000，　　F（010）= 000，　　F（100）= 000

F（011）= 111，　F（101）= 111，　　F（110）= 111，　　F（111）= 111

计算平均错误概率 p_e，由于信源等概分布，有 $q(\boldsymbol{x}_1)=q(\boldsymbol{x}_2)=\frac{1}{2}$。

根据式（5-4）得

$$p_e=\sum_{\boldsymbol{x}\neq\boldsymbol{x}_k}\sum_{\boldsymbol{y}}p(\boldsymbol{xy})=\sum_{\boldsymbol{x}\neq\boldsymbol{x}_k}\sum_{\boldsymbol{y}}p(\boldsymbol{y}|\boldsymbol{x})\,q(\boldsymbol{x})$$

$$=\frac{1}{2}\times 2p^3+\frac{1}{2}\times 6(1-p)p^2=3p^2-2p^3$$

与式（5-1）的计算结果一致。

5.3　信道编码定理

信号在传输过程中不可避免地会拾取干扰，以致产生传输错误，选择合适的编译码规则可以使错误概率尽可能小，信道编码定理（香农第二定理）指出：信道容量 C 是满足错误概率 $p_e\to 0$ 时，信道所能容纳的信息传输率的极限值。

定理 5.1　对于任何离散无记忆信道 DMC，存在信息传输率为 R，长为 n 的码，当 $n\to\infty$ 时，平均差错概率 $p_e<\exp\{-nE(R)\}\to 0$，式中 $E(R)$ 为可靠性函数，$E(R)$在 $0<R<C$ 的范围内为正。

定理 5.1 说明：信道容量 C 是保证无差错传输时，信息传输率 R 的极限值，对于固定信道，C 值是一定的，它是衡量信道质量的一个重要物理量。

上述定理也称**有噪信道编码定理**，即 Shannon 第二定理。

下面进行证明。

1．随机编码方法

随机编码方法是 Shannon 引入的一个数学方法，它是信息论的基本技巧，通过这一方法证得了有噪信道编码定理——Shannon 第二定理。

信道输入字符集 $A=\{a_1,a_2,\cdots,a_k\}$，共 k 个字符。

选用长为 n 的定长码，共可构成 k^n 个矢量，设有 M 个消息待传输（$M<k^n$），每次随机地从 k^n 个矢量中抽出 M 个矢量构成一个码集 $\boldsymbol{C}$（允许重复取），有

$$\boldsymbol{C}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m,\cdots,\boldsymbol{x}_M\}$$

共可构成 k^{nM} 个这样的码集。

由随机编码的构造方法知道，得到任何一个码字的概率相同，且相互独立。

在 5.2 节中，导出了按最大后验概率译码准则译码时的平均错误概率 p_e 的计算式（5-4），但当信道中传输的码字的码长 n 很大时，p_e 的计算变得很困难，当 n>100 时，用计算机精确地计算出 p_e 也很困难，所以转为寻找 p_e 的下界和上界，用它们来对 p_e 进行估值。

但下界的估算更为复杂，这里只讨论上界，给出常用的 Gallager 上界。

2．Gallager上界

因为上述按随机编码方法构成的码集是等概分布的，按照最大释然译码准则译码，在发送码矢 $\boldsymbol{x}_k$ 时，要得到正确译码须满足

$$p(\boldsymbol{y}|\boldsymbol{x}_k) > p(\boldsymbol{y}|\boldsymbol{x}_m) \quad \forall m \neq k, \quad \text{即} \frac{p(\boldsymbol{y}|\boldsymbol{x}_k)}{p(\boldsymbol{y}|\boldsymbol{x}_m)} > 1$$

则

$$\left[\frac{p(\boldsymbol{y}|\boldsymbol{x}_k)}{p(\boldsymbol{y}|\boldsymbol{x}_m)}\right]^{\lambda} > 1 \quad \lambda > 0 \tag{5-5}$$

定义示性函数

$$I_k(\boldsymbol{y}) = \begin{cases} 0 & p(\boldsymbol{y}|\boldsymbol{x}_k) > p(\boldsymbol{y}|\boldsymbol{x}_m) \\ 1 & p(\boldsymbol{y}|\boldsymbol{x}_k) \leq p(\boldsymbol{y}|\boldsymbol{x}_m) \end{cases} \quad m = 1, 2, \cdots, M$$

则发送码矢 $\boldsymbol{x}_k$ 判错的概率为

$$p_e(\boldsymbol{x}_k) = \sum_{\boldsymbol{y}} I_k(\boldsymbol{y}) p(\boldsymbol{y}|\boldsymbol{x}_k) \tag{5-6}$$

根据式（5-5），当 $0 \leqslant \lambda, \rho \leqslant 1$ 时，下式成立

$$\left\{\sum_{m \neq k}\left[\frac{p(\boldsymbol{y}|\boldsymbol{x}_m)}{p(\boldsymbol{y}|\boldsymbol{x}_k)}\right]^{\lambda}\right\}^{\rho} \geq \begin{cases} 0 & \text{当 } p(\boldsymbol{y}|\boldsymbol{x}_k) > p(\boldsymbol{y}|\boldsymbol{x}_m) \\ 1 & \text{当 } p(\boldsymbol{y}|\boldsymbol{x}_k) \leq p(\boldsymbol{y}|\boldsymbol{x}_m) \end{cases}$$

用示性函数 $I_k(\boldsymbol{y})$表示上式，即

$$I_k(\boldsymbol{y}) \leq \left\{\sum_{m \neq k}\left[\frac{p(\boldsymbol{y}|\boldsymbol{x}_m)}{p(\boldsymbol{y}|\boldsymbol{x}_k)}\right]^{\lambda}\right\}^{\rho} \tag{5-7}$$

将式（5-7）代入式（5-6），得

$$p_e(\boldsymbol{x}_k) \leq \sum_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}_k)\left\{\sum_{m \neq k}\left[\frac{p(\boldsymbol{y}|\boldsymbol{x}_m)}{p(\boldsymbol{y}|\boldsymbol{x}_k)}\right]^{\lambda}\right\}^{\rho}$$

式中，$0 \leqslant \lambda, \rho \leqslant 1$，因为$\lambda, \rho$ 都是任意数，可取 $\lambda = \dfrac{1}{1+\rho}$，则有

$$p_e(\boldsymbol{x}_k) \leq \sum_{\boldsymbol{Y}} p(\boldsymbol{y}|\boldsymbol{x}_k)^{\frac{1}{1+\rho}}\left\{\sum_{m \neq k}\left[p(\boldsymbol{y}|\boldsymbol{x}_m)\right]^{\frac{1}{1+\rho}}\right\}^{\rho} \tag{5-8}$$

式（5-8）称为 Gallager 上界。

3．随机编码错误概率上界

Gallager 限仅给出发送码矢 $\boldsymbol{x}_k$ 时的错误概率上界，还要对全部码失求平均，下面对上述的随机编码集合求平均。

对于随机编码，各码字等概且独立，有 $q(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m,\cdots,\boldsymbol{x}_M)=\prod_{m=1}^{M}q(\boldsymbol{x}_m)$，对式（5-8）求统计平均值，得平均错误概率上界

$$
\begin{aligned}
p_{\mathrm{e}} &= \sum_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M} q(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m,\cdots,\boldsymbol{x}_M)p_{\mathrm{e}}(\boldsymbol{x}_k) \\
&\le \sum_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M} q(\boldsymbol{x}_1)q(\boldsymbol{x}_2)\cdots q(\boldsymbol{x}_M)\sum_{\boldsymbol{y}} p(\boldsymbol{y}|\boldsymbol{x}_k)^{\frac{1}{1+\rho}}\left\{\sum_{m\neq k}\left[p(\boldsymbol{y}|\boldsymbol{x}_m)\right]^{\frac{1}{1+\rho}}\right\}^{\rho} \\
&= \sum_{\boldsymbol{y}}\sum_{\boldsymbol{x}_k} q(\boldsymbol{x}_k)\,p(\boldsymbol{y}|\boldsymbol{x}_k)^{\frac{1}{1+\rho}}\sum_{\boldsymbol{x}_m} q(\boldsymbol{x}_m)\left[\sum_{m\neq k} p(\boldsymbol{y}|\boldsymbol{x}_m)^{\frac{1}{1+\rho}}\right]^{\rho} \qquad (5\text{-}9)
\end{aligned}
$$

先看后面一项 $\sum_{\boldsymbol{x}_m} q(\boldsymbol{x}_m)\left[\sum_{m\neq k} p(\boldsymbol{y}|\boldsymbol{x}_m)^{\frac{1}{1+\rho}}\right]^{\rho}$，因为 0≤ρ≤1，而 x^{ρ}是 x 的∩型凸函数，由∩型凸函数的定义知，函数均值小于等于均值函数，即有

$$
\sum_{\boldsymbol{x}_m} q(\boldsymbol{x}_m)\left[\sum_{m\neq k} p(\boldsymbol{y}|\boldsymbol{x}_m)^{\frac{1}{1+\rho}}\right]^{\rho} \le \left[\sum_{\boldsymbol{x}_m} q(\boldsymbol{x}_m)\sum_{m\neq k} p(\boldsymbol{y}|\boldsymbol{x}_m)^{\frac{1}{1+\rho}}\right]^{\rho} \qquad (5\text{-}10)
$$

因为$\forall\boldsymbol{x}_m$（$m=1,2,\cdots,M$）都通过同一信道传输，故 $p(\boldsymbol{y}|\boldsymbol{x}_m)$值相同，记为 $p(\boldsymbol{y}|\boldsymbol{x})$，代入式（5-10）得

$$
\sum_{\boldsymbol{x}_m} q(\boldsymbol{x}_m)\left[\sum_{m\neq k} p(\boldsymbol{y}|\boldsymbol{x}_m)^{\frac{1}{1+\rho}}\right]^{\rho} \le \left[(M-1)\sum_{\boldsymbol{x}} q(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\right]^{\rho} \qquad (5\text{-}11)
$$

将式（5-11）代回式（5-9），得

$$
\begin{aligned}
p_{\mathrm{e}} &\le \sum_{\boldsymbol{y}}\sum_{\boldsymbol{x}_k} q(\boldsymbol{x}_k)\,p(\boldsymbol{y}|\boldsymbol{x}_k)^{\frac{1}{1+\rho}}\left[(M-1)\sum_{\boldsymbol{x}} q(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\right]^{\rho} \\
&= (M-1)^{\rho}\sum_{\boldsymbol{y}}\left[\sum_{\boldsymbol{x}} q(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\right]^{\rho+1} \qquad (5\text{-}12)
\end{aligned}
$$

这是平均错误概率的一个上界，在这个上界的推导过程中，并没有强调离散无记忆信道，因此它对有记忆的离散输入、输出及连续输出信道都适用，只不过在连续情况下，$p(\boldsymbol{y}|\boldsymbol{x})$ 表示概率密度。

4．离散无记忆信道DMC的错误概率上界

对于 n 维离散无记忆信道 DMC，有 $p(\boldsymbol{y}|\boldsymbol{x})=\prod_{i=1}^{n}p(y_i|x_i)$，对于随机编码 $q(\boldsymbol{x})=\prod_{i=1}^{n}q(x_i)$（信源也离散无记忆），将这两个关系式代入式（5-12），并记 DMC 的平均错误概率为 $\overline{p_{\mathrm{e}}}$，有

$$
\overline{p_{\mathrm{e}}} \le (M-1)^{\rho}\sum_{y_1}\cdots\sum_{y_N}\left[\sum_{x_1}\cdots\sum_{x_N} q(x_1)p(y_1|x_1)^{\frac{1}{1+\rho}}\cdots q(x_N)p(y_N|x_N)^{\frac{1}{1+\rho}}\right]^{1+\rho}
$$

$$=(M-1)^{\rho}\prod_{i=1}^{n}\sum_{y_i}\left[\sum_{x_i}q(x_i)p(y_i|x_i)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{5-13}$$

因为序列 $\boldsymbol{x}=x_1,\cdots,x_i,,\cdots,x_n$ 中的$\forall x_i$，都取自同一符号集 $A=\{a_1,a_2,\cdots,a_k\}$，故分布 $q(x_i)$相同, $\forall x_i$，又都通过同一信道传输，故 $p(y_i|x_i)$也相同，记

$$\sum_{y_i}\left[\sum_{x_i}q(x_i)p(y_i|x_i)^{\frac{1}{1+\rho}}\right]^{1+\rho}=\sum_{y}\left[\sum_{x}q(x)p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{5-14}$$

将式（5-14）代入式（5-13），得

$$\overline{p_{\mathrm{e}}}\leq (M-1)^{\rho}\left\{\sum_{y}\left[\sum_{x}q(x)p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{n}$$

$$<M^{\rho}\left\{\sum_{y}\left[\sum_{x}q(x)p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{n} \tag{5-15}$$

5．可靠性函数$E(R)$

信息传输率 $R=\dfrac{\ln M}{n}\ \Rightarrow\ M=\mathrm{e}^{Rn}$，则式（5-15）可写为

$$\overline{p_{\mathrm{e}}}<\mathrm{e}^{Rn\rho}\mathrm{e}^{\ln\left\{\sum_{y}\left[\sum_{x}q(x)p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{n}} \tag{5-16}$$

记

$$-\ln\sum_{y}\left[\sum_{x}q(x)p(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}=E_0(\rho,q),\quad 0\leq\rho\leq 1$$

则式（5-16）可写成

$$\overline{p_{\mathrm{e}}}<\exp\left\{-n\left[-\rho R+E_0(\rho,q)\right]\right\} \tag{5-17}$$

为了使上界更紧致，对式（5-17）右边求极小值，即对指数$-\rho R+E_0(\rho,q)$求极大值，记这个极大值为

$$E(R)\triangleq\max_{0\leq\rho\leq 1}\max_{q}\left[E_0(\rho,q)-\rho R\right]$$

则式（5-17）可写为

$$p_{\mathrm{e}}<\exp\{-nE(R)\}$$

称 $E(R)$为可靠性函数，也称随机编码指数，它与信道转移概率 $p(y|x)$有关。

可以证明，在 $0\leqslant R\leqslant C$ 的范围内，$E(R)$是下降的、下凹的正值函数，说明 $E(R)$有界，这样，当 $\lim\limits_{n\to\infty}nE(R)\to\infty$ 时，有 $\exp\{-nE(R)\}\to 0$，从而 $p_{\mathrm{e}}<\exp\{-nE(R)\}\to 0$。

这就证明了信道编码定理。 证毕

可靠性函数 $E(R)$在信道编码中有重要意义，它表示了在码长 n 一定的条件下，最佳编码平均错误概率的一个上界，同时也说明了 p_{e}随 $n\to\infty$而趋于 0 的速率，在规定了 p_{e}值后，$E(R)$可帮助选择合适的码长 n 和信息传输率 R。

5.4 费诺引理及信道编码逆定理

信道编码定理指出，当信息传输率 $R<C$ 时，一定存在着错误概率 $p_e \to 0$ 的好码。本节介绍的信道编码逆定理则指出，若 $R>C$，则无论采用什么方法，都不能使 $p_e \to 0$。

信道编码逆定理的证明要用到费诺（Fano）不等式，它是信息论中一个著名的不等式，下面予以介绍。

5.4.1 费诺不等式

设信道输入符号 X 和输出符号 Y 取自同一符号集 $A=\{a_1, a_2, \cdots, a_k\}$，则传输过程中的错误概率 p_e 和信道疑义度 $H(X|Y)$ 之间满足下列关系式：

$$H(X|Y) \leqslant H_2(p_e) + p_e \log(k-1) \tag{5-18}$$

式（5-18）就是著名的 **Fano 不等式**。

证明：设信道输入符号为 x_i，输出符号为 y_j，根据最大后验概率准则进行判决，则发 x_i 收到 y_j 判对的概率为 $\phi(x_j|y_j)$，判错的概率为

$$p(e|y_j) = 1 - \phi(x_j|y_j) = \sum_{i \neq j} \phi(x_i|y_j) \tag{5-19}$$

另一方面，我们可算出条件熵

$$\begin{aligned} H(X|y_j) &= -\sum_i \phi(x_i|y_j) \log \phi(x_i|y_j) \\ &= -\phi(x_j|y_j) \log \phi(x_j|y_j) - \sum_{i \neq j} \phi(x_i|y_j) \log \phi(x_i|y_j) \end{aligned} \tag{5-20}$$

将式（5-20）分两项考虑。

第一项，根据式（5-19）可得

$$-\phi(x_j|y_j) \log \phi(x_j|y_j) = -[1-p(e|y_j)] \log [1-p(e|y_j)] \tag{5-21}$$

第二项，

$$\begin{aligned} &-\sum_{i \neq j} \phi(x_i|y_j) \log \phi(x_i|y_j) \\ &= -\log p(e|y_j) \sum_{i \neq j} \phi(x_i|y_j) - p(e|y_j) \sum_{i \neq j} \left[\frac{\phi(x_i|y_j)}{p(e|y_j)} \log \frac{\phi(x_i|y_j)}{p(e|y_j)} \right] \\ &\leqslant p(e|y_j) \log p(e|y_j) + p(e|y_j) \log(k-1) \end{aligned} \tag{5-22}$$

式中，$\left[-\dfrac{\phi(x_i|y_j)}{p(e|y_j)} \log \dfrac{\phi(x_i|y_j)}{p(e|y_j)} \right]$ 是（$k-1$）个事件的熵，根据极大熵定理，它不大于这（$k-1$）个事件等概分布时的熵 $\log(k-1)$。

将式（5-21）和式（5-22）代回到式（5-20），得

$$\begin{aligned} H(X|y_j) &\leqslant -[1-p(e|y_j)] \log [1-p(e|y_j)] - p(e|y_j) \log p(e|y_j) + p(e|y_j) \log(k-1) \\ &= H_2[p(e|y_j)] + p(e|y_j) \log(k-1) \end{aligned} \tag{5-23}$$

设 Y 空间分布为 $\{\omega(y_j), j=1, 2, \cdots, k\}$，对式（5-23）两边关于 Y 空间求和，得

$$H\left(X\middle|Y\right)=\sum_j \omega(y_j)H\left(X\middle|y_j\right)$$

$$\leq \sum_j \omega(y_j)H_2\left[p(e\middle|y_j)\right]+\sum_j \omega(y_j)p(e\middle|y_j)\log(k-1)$$

$$=H_2(p_e)+p_e\log(k-1)$$

这就得到了 Fano 不等式（5-18）：$H(X|Y)\leqslant H_2(p_e)+p_e\log(k-1)$

- Fano 不等式的物理意义：

进行一次判决后，关于 X 的疑义度可分成两项：

（1）是否判对，疑义度为 $H_2(p_e)$；

（2）如果判决出错（概率为 p_e），错在 $k-1$ 个符号中的哪一个？疑义度不会超过 $\log(k-1)$。

- 将 Fano 不等式推广到 L 维矢量情况：

设 $\boldsymbol{x}=x_1,\cdots,x_l,\ \cdots,x_L$，$\boldsymbol{y}=y_1,\ \cdots,y_l,\ \cdots,y_L$，皆为 L 维矢量，$\forall x_l,y_l\in A=\{a_1,a_2,\ \cdots,a_k\}$，记 p_e 为错误概率，则

$$H(\boldsymbol{X}|\boldsymbol{Y})\leqslant L\left[H_2(p_e)+p_e\log(k-1)\right] \tag{5-24}$$

证明：根据熵的链规则式（2-44），有

$$H(\boldsymbol{X}|\boldsymbol{Y})=H(X_1|\boldsymbol{Y})+H(X_2|\boldsymbol{Y}X_1)+\ \cdots\ +H(X_L|\boldsymbol{Y}\boldsymbol{X}_1^{L-1})$$

$$\leqslant H(X_1|Y_1)+H(X_2|Y_2)+\ \cdots\ +H(X_L|Y_L) \qquad \text{（条件熵≤无条件熵）}$$

$$=\sum_{l=1}^{L}H(X_l\,|\,Y_l)\leq\ \sum_{l=1}^{L}[H_2(p_e)+p_e\log(k-1)] \qquad \text{（Fano 不等式）}$$

$$=L\left[H_2(p_e)+p_e\log(k-1)\right] \qquad \text{（}\forall x_l,y_l\text{ 都取自同一符号集）}$$

证毕

5.4.2 信道编码逆定理

在讲述信道编码逆定理以前，先介绍一个引理。

引理 5.1 设 $\boldsymbol{u}=u_1,\ \cdots,u_l,\ \cdots,u_L$ 为 L 维随机矢量，$\forall u_l$ 取自同一符号集$\{U\}$，则

$$\lim_{L\to\infty}\frac{H(\boldsymbol{U})}{L}=H(U) \tag{5-25}$$

式中，$H(\boldsymbol{U})$表示 L 维熵，$H(U)$表示一维熵。

$$H(\boldsymbol{U})=H(\boldsymbol{U}_1^{L-1},U_L)=H(\boldsymbol{U}_1^{L-1})+H(U_L|\boldsymbol{U}_1^{L-1})$$

$$\frac{H(\boldsymbol{U})}{L}=\frac{H(\boldsymbol{U}_1^{L-1})}{L}+\frac{H(U_L\left|\boldsymbol{U}_1^{L-1}\right.)}{L} \tag{5-26}$$

另一方面，根据熵的链规则，有

$$H(\boldsymbol{U})=H(U_1)+H(U_2|U_1)+H(U_3|U_1U_2)+\ \cdots\ +H(U_L|U_1U_2\cdots U_{L-1})$$

$$\geqslant H(U_1|U_1U_2\cdots U_{L-1})+H(U_2|U_1U_2\cdots U_{L-1})+\ \cdots\ +H(U_L|U_1U_2\cdots U_{L-1})$$

$$=L\,H(U_L|\boldsymbol{U}_1^{L-1}) \qquad \text{（因 }\forall u_l\text{ 取自同一符号集）}$$

式中，$\boldsymbol{U}_1^{L-1}=U_1U_2\cdots U_{L-1}$。故

$$H(U_L\left|\boldsymbol{U}_1^{L-1}\right.)\leq\ \frac{H(\boldsymbol{U})}{L} \tag{5-27}$$

将式（5-27）代入式（5-26），得

$$\frac{H(\boldsymbol{U})}{L} \leq \frac{H(\boldsymbol{U}_1^{L-1})}{L} + \frac{1}{L}\frac{H(\boldsymbol{U})}{L}$$

即
$$\frac{H(\boldsymbol{U})}{L} \leq \frac{H(\boldsymbol{U}_1^{L-1})}{L-1} \tag{5-28}$$

式（5-28）表示$\left\{\frac{H(\boldsymbol{U})}{L}\right\}$是 L 的单调降序列。而由熵的非负性，说明$\frac{H(\boldsymbol{U})}{L}$有下界，其下界为 $H(U)$，即

$$\lim_{L\to\infty}\frac{H(\boldsymbol{U})}{L} = H(U)$$

考虑图 5-6 所示的通信过程，$\boldsymbol{u}$为信源输出的 L 维序列，$\boldsymbol{x}$为经信道编码器编码后输出的 n 长码字，$\boldsymbol{y}$是 $\boldsymbol{x}$经 n 维有扰扩展信道传输后的输出，$\hat{\boldsymbol{u}}$是对 $\boldsymbol{y}$的估值。矢量 $\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{u}}$ 中的所有元素都取自同一符号集 $A = \{a_1, a_2, \cdots, a_k\}$。

根据数据处理定理：不论经过何种数据处理，信息量均减少。则由图 5-6 可见 $I(\boldsymbol{U};\hat{\boldsymbol{U}}) \leq I(\boldsymbol{X};\boldsymbol{Y})$，对于离散无记忆信道，根据定理 4.1 有 $C_n \leqslant nC$，则

$$I(\boldsymbol{U};\hat{\boldsymbol{U}}) \leq I(\boldsymbol{X};\boldsymbol{Y}) \leqslant C_n \leqslant nC \tag{5-29}$$

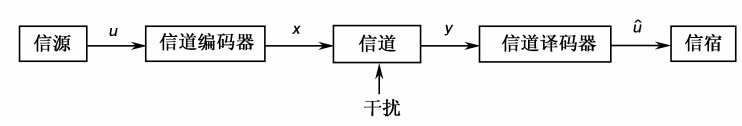


图 5-6　通信过程

另一方面，
$$I(\boldsymbol{U};\hat{\boldsymbol{U}}) = H(\boldsymbol{U}) - H(\boldsymbol{U}\mid\hat{\boldsymbol{U}})$$

$$H(\boldsymbol{U}) - I(\boldsymbol{U};\hat{\boldsymbol{U}}) = H(\boldsymbol{U}\mid\hat{\boldsymbol{U}}) \tag{5-30}$$

将 Fano 不等式的矢量式（5-24）及式（5-29）代入式（5-30）得

$$H(\boldsymbol{U}) - nC \leqslant L\,[H_2(p_e) + p_e \log(k-1)] \tag{5-31}$$

考虑到式（5-25），在 $L\to\infty$条件下，$H(\boldsymbol{U}) = L\,H(U)$，代入到式（5-31）并整理得

$$H_2(p_e) + p_e \log(k-1) \geqslant H(U) - \frac{nC}{L} = \frac{n}{L}\left[\frac{L}{n}H(U) - C\right] = \frac{n}{L}\left[R_D - C\right]$$

即

$$H_2(p_e) + p_e \log(k-1) \geq \frac{n}{L}\left[R_D - C\right] \tag{5-32}$$

式中，$R_D = \frac{L}{n}H(U)$为信息传输率。

从式（5-32）可看出，当 R_D＞C 时，右边为大于零的数，说明差错概率有确定下界，不能以任何方式趋于零，这就是下面的信道编码逆定理。

定理 5.2　信道编码逆定理　信道容量 C 是可靠通信系统信息传输率的上界，当 R＞C 时，不可能存在任何方法使差错概率任意小。

本 章 小 结

码矢集合 $\boldsymbol{C}=\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$，码矢 $\boldsymbol{x}_k \in \boldsymbol{C}$通过信道转移概率为 $p(\boldsymbol{y} \mid \boldsymbol{x}_k)$的信道传输，输出矢量 $\boldsymbol{y}$，信道译码器估值为 $\hat{\boldsymbol{x}}_k = F(\boldsymbol{y})$，$\hat{\boldsymbol{x}}_k \in \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$。

最大后验概率译码准则 $\quad \phi(\hat{\boldsymbol{x}}_k \mid \boldsymbol{y}) > \phi(\boldsymbol{x}_m \mid \boldsymbol{y}) \begin{cases} k \neq m \quad m=[1,2,\cdots,M] \\ x_m \in \{\boldsymbol{C}\} \end{cases}$

极大似然译码准则 $\quad p(\boldsymbol{y} \mid \hat{\boldsymbol{x}}_k) > p(\boldsymbol{y} \mid \boldsymbol{x}_m) \begin{cases} k \neq m \quad m=[1,2,\cdots,M] \\ x_m \in \{\boldsymbol{C}\} \end{cases}$

平均译码错误概率 $\quad p_{\mathrm{e}} = \sum_{\boldsymbol{x}-\boldsymbol{x}_k} \sum_{\boldsymbol{y}} p(\boldsymbol{x}\boldsymbol{y})$

信道编码定理

对于任何离散无记忆信道 DMC，存在信息传输率为 R，长为 n 的码，当 $n \to \infty$ 时，平均差错概率 $p_{\mathrm{e}} < \exp\{-nE(R)\} \to 0$，式中，$E(R)$ 为可靠性函数，$E(R)$在 $0<R<C$ 的范围内为正。

信道编码逆定理

信道容量 C 是可靠通信系统信息传输率的上界，当 $R>C$ 时，不可能存在任何方法使差错概率任意小。

Fano 不等式 $\quad H(X \mid Y) \leqslant H_2(p_{\mathrm{e}}) + p_{\mathrm{e}} \log(k-1)$

思考题与习题

5.1 R 为信息传输率，根据 Shannon 第二定理，当码长 $n \to \infty$ 时，满足什么关系式，可使错误概率 $p_{\mathrm{e}} \to 0$?

5.2 写出 Fano 不等式，其中哪一项表示是否判对的疑义度？$\log(k-1)$又表示什么？

5.3 根据 Shannon 定理，说明哪个物理量是保证无差错传输时信息传输率 R 的上限值，哪个物理量又是信源可压缩信息的最低极限。

5.4 最大后验概率译码准则就是最小错误译码准则，对吗？

5.5 极大似然译码准则就是最小错误译码准则，对吗？

5.6 在信源等概分布时，极大似然译码准则就是最小错误译码准则，对吗？

5.7 信道编码以增加冗余度来提高可靠性，这种说法对吗？

5.8 离散无记忆信源，熵为 $H(X)$，对信源的 L 长序列进行等长编码，码字是长为 n 的 D 进制符号串，试问满足什么条件时，可实现无失真编码？

5.9 二进制删除信道的信道转移概率矩阵位 $\boldsymbol{P} = \begin{bmatrix} 1-\delta & \delta & 0 \\ 0 & \delta & 1-\delta \end{bmatrix}$，证明对于此信道来说，极大似然译码准则就等于最小错误译码准则。

5.10 离散无记忆信道 DMC，转移概率矩阵为 $\boldsymbol{P} = \begin{bmatrix} 1/2 & 1/6 & 1/3 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 1/3 & 1/2 \end{bmatrix}$

（1）$q(x_1)=\frac{1}{2}, q(x_2)=\frac{1}{4}, q(x_3)=\frac{1}{4}$，求最佳判决译码及误码率 p_{e1};

（2）若信源等概分布，求最佳判决译码及误码率 p_{e2}。

5.11　设有一离散无记忆信道，其信道转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}1/2 & 1/3 & 1/6\\ 1/3 & 1/6 & 1/2\\ 1/6 & 1/2 & 1/3\end{bmatrix}$，若信源分布为 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3\\ 2/7 & 3/7 & 2/7\end{bmatrix}$，按最大后验概率译码，求误码率 p_e。

5.12　一离散无记忆信道的信道转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}1/2 & 1/2 & 0 & 0 & 0\\ 0 & 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/2 & 1/2 & 0\\ 0 & 0 & 0 & 1/2 & 1/2\\ 1/2 & 0 & 0 & 0 & 1/2\end{bmatrix}$，

（1）找出一个码长为 2 的重复码（含 5 个码字），其信息传输率为 $\frac{1}{2}\log 5$，按极大似然译码准则设计译码器，若码字等概输入译码器，求译码器输出端的平均错误概率 p_e。

（2）找出一个码长为 2 的码，使平均错误概率 $p_e=0$。

5.13　信道输入 $\{x_0, x_1\}$，输出 $\{y_0, y_1, y_2\}$，信道转移概率 $\boldsymbol{P}=\begin{bmatrix}1/2 & 1/2 & 0\\ 0 & 1/2 & 1/2\end{bmatrix}$，求

（1）使 $I(X;Y)$ 达到信道容量的最佳输入分布;

（2）最佳分布下的 $E_0(\rho, q)$ 函数。

5.14　计算下述信道的 $E_0(1,q)$，$E_0(1)=\max\limits_{q} E_0(1,q)$。

（1）二进制对称信道，固有误码率为 p;

（2）二元 Z 信道，转移概率为 p。

5.15　符号取自 $\{0, 1\}$ 的长度为 3 的序列，共有 2^3 个。信源含 8 个消息，对其编码，码字从 2^3 个长度为 3 的序列中独立、等概选取，若采用极大似然译码规则，分别求通过以下三种信道传输时的平均错误概率 p_e。

（1）二进制对称信道，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$;

（2）二进制 Z 信道，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1 & 0\\ p & 1-p\end{bmatrix}$;

（3）二进制删除信道，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}1-p & p & 0\\ 0 & p & 1-p\end{bmatrix}$。

5.16　信源 $\begin{bmatrix}X\\ q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3\\ 1/3 & 1/3 & 1/3\end{bmatrix}$，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.4 & 0.3 & 0.3\\ 0.1 & 0.9 & 0\end{bmatrix}$，信道输出符号 $Y=\{y_1, y_2, y_3\}$。

（1）计算信源熵 $H(X)$和噪声熵 $H(Y|X)$;

（2）计算从发送端观察到的平均错误概率 p_{ex};

（3）计算从接收端观察到的平均错误概率 p_{ey};

（4）你能否通过信道转移概率矩阵，对该信道作一个好坏评价？

5.17　信源符号$\{x_0 \quad x_1 \quad x_2 \quad x_3\}$等概分布，取码长为 4，对其进行二进制等长编码，编码为$\begin{cases} x_0:1111 \\ x_1:0000 \\ x_2:1100 \\ x_3:0011 \end{cases}$，信道为二进制对称信道，固有误码率为 p（$p<0.5$），试选择一种译码规则，使平均错误概率 p_e 最小。

5.18　信源符号集$\{x_0, x_1, x_2, x_3, x_4\}$，等概分布，给定一离散无记忆信道，信道转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{bmatrix}$

（1）计算信道容量 C;

（2）设计一码长为 2 的重复码，按极大似然译码准则译码，求平均错误概率 p_e。

5.19　信道输入符号集 $X = \{0, e, 1\}$，输出符号集 $Y = \{0, 1\}$，信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}$，信源有 4 个消息$\{u_0, u_1, u_2, u_3\}$，编码为$\begin{cases} u_0:00ee \\ u_1:01ee \\ u_2:10ee \\ u_3:11ee \end{cases}$，接收矢量 $\boldsymbol{y}=y_0\,y_1\,y_2\,y_3$，译码规则 $F(y_0\,y_1\,y_2\,y_3)=y_0\,y_1\,e\,e$。

（1）求信息传输率 R_D；

（2）证明在上述译码规则下，对所有码字错误概率等于零。

第 6 章
率失真编码

内容提要

数据压缩是信息传输和处理的重要研究内容，为提高传输效率，在允许的情况下可对信源进行压缩编码，率失真理论研究的就是在允许一定失真的前提下，对信源的压缩编码，这一理论已广泛应用于语音和图像的信息压缩技术。率失真信源编码定理（香农第三定理）指出：率失真函数 $R(D)$ 就是在给定失真测度条件下，对信源熵可压缩的最低程度。

本章只限于研究率失真理论最基本的内容：失真测度、率失真函数、率失真函数的定义域、值域、性质及定量计算。$R(D)$ 的计算很烦琐，文中通过两个例子介绍了几种特殊情况下 $R(D)$ 的求法，一般情况下只能用参数法求解。

知识要点

失真测度，平均失真，率失真函数，率失真编码定理。

教学建议

率失真函数是信息论中一个重要的物理量，从它的定义 $R(D)\triangleq\min_{p(y|x)}\{I(X;Y):\ \overline{D}\leq D\}$ 看，即在满足平均失真小于等于给定失真 D 的前提下，在各种可能的信道（其特性用信道转移概率矩阵 $p(y|x)$ 描述）中，求平均互信息量的最小值。这里所指的各种可能信道，并不是真正的信道，只是实验信道，它对应不同的信源编码方法，读者要搞清楚这个概念。率失真编码定理是香农第三定理，是信息论中的重要定理，证明过程冗长，本文省略了其证明过程，请读者弄清楚其物理意义。建议学时数为 6 学时。

6.1 失真测度与平均失真

第 3 章研究了无失真信源编码，要求在信源无失真前提下，以最高的信息传输率来传递信息，而在接收端应以任意小的失真恢复原信号。而定理 5.1（Shannon 第二定理）指出，只有在信息传输率 R 小于信道容量 C 时，才可能找到一种编码，以任意小的错误概率实现几乎无差错传输；反之，若 $R>C$，则在任何信道中都不可能实现无差错传输。而对每个信道而言，其信道容量 C 是一个定值，即信息传输率有一个极限值，这就出现了矛盾。

随着科学技术的发展，高科技、高产业、信息全球化，这些都离不开对大流量信息的处理、存储、传输，对信号再现要求越趋完美，对信息处理、存储、传输的要求就越高，这些都需要花费大量的软、硬件开销。为了提高信号传输系统软、硬件的使用效率，在一定保真度前提下，对信源进行压缩是信息传输和处理中的重要研究课题。在实际情况中，信息接收方并不要求完全无差错地再现原信号，而允许接收信号有一定的失真，这在很多情况下是可以接受的。例如，人们接听电话时，由于人耳对音频信号的带宽及分辨率有限，大部分人只能分辩几千赫兹到十几千赫兹的音频，即使语音信号有一些失真，人们也能听明白电话的内容。又如电视播放，从理论上说要将一个连续动作完全无误地反映出来，应该用无穷多个静态画面连续播放，否则将会有动作不连贯的感觉，从而产生失真。而实际上利用人们的“视觉滞留”现象，只要达到 25 帧/秒的传输率，就能满足人们观看电视的要求，所以失真没有必要完全消除。

在允许一定失真的前提下，从提高传输效率的角度出发，可以对信源信息量事先进行压缩再予传输，那么可以压缩到什么程度呢？本章要讨论的问题就是给定一个失真度，求出在平均失真小于给定值的条件下，信源所能压缩的最低程度，即率失真函数 $R(D)$。

1．失真测度 $d(x,y)$

给定离散信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_I \\ q(x_1) & q(x_2) & \cdots & q(x_I) \end{bmatrix}$，信源的输出即信道的输入，而信道输出符号为 $Y=\{y_1, y_2, \cdots, y_J\}$，信源发送符号 x_i，而信道输出符号 y_j 引起的失真用 $d(x_i, y_j)$（$i=1, 2, \cdots, I$; $j=1, 2, \cdots, J$）表示，简记为 d_{ij}，将所有的 d_{ij} 列出来，可以得到下面的失真测度矩阵

$$[d]=\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1J} \\ d_{21} & d_{22} & \cdots & d_{2J} \\ \vdots & \vdots & \cdots & \vdots \\ d_{I1} & d_{I2} & \cdots & d_{IJ} \end{bmatrix} \tag{6-1}$$

【例 6.1】 汉明（Hamming）失真测度

信源输出符号 $X=\{x_1, x_2, \cdots, x_K\}$，信道输出符号 $Y=\{y_1, y_2, \cdots, y_K\}$，约定失真测度

$$\begin{cases} y_i = x_i & \text{无误码} \quad d_{ii}=0 \\ y_j = x_i (i \neq j) & \text{误码} \quad d_{ij}=1 \end{cases} \qquad i, j = 1, 2, \cdots, K$$

上述约定可以用矩阵表示为

$$[d]=\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & 0 & \cdots & 1\\ \vdots & \vdots & \cdots & \vdots\\ 1 & 1 & \cdots & 0\end{bmatrix}$$

式中，$d_{ij}\geqslant 0, i,j=1,2,\cdots; K$ 为信源方发送符号 x_i 而信宿方判为 y_j 引起的失真度。

【例 6.2】 平方误差失真测度

信源输出符号 $X=\{0,1,2\}$，信道输出符号 $Y=\{0,1,2\}$，给出失真测度

$$d_{ij}=(x_i-y_j)^2 \qquad i,j=0,1,2$$

则失真测度矩阵为

$$[d]=\begin{bmatrix}0 & 1 & 4\\ 1 & 0 & 1\\ 4 & 1 & 0\end{bmatrix}$$

【例 6.3】 绝对值误差失真测度

信源输出符号 $X=\{0,1,2\}$，信道输出符号 $Y=\{0,1,2\}$，给出失真测度

$$d_{ij}=|x_i-y_j| \qquad i,j=0,1,2$$

则失真测度矩阵为

$$[d]=\begin{bmatrix}0 & 1 & 2\\ 1 & 0 & 1\\ 2 & 1 & 0\end{bmatrix}$$

对于矢量传输情况，若信道的输入、输出均为 N 长序列 $\boldsymbol{X}=X_1X_2\cdots X_N$, $\boldsymbol{Y}=Y_1Y_2\cdots Y_N$，定义失真测度为

$$d^{(N)}(\boldsymbol{X},\boldsymbol{Y})=\frac{1}{N}\sum_{k=1}^{N}d(X_k,Y_k) \tag{6-2}$$

【例 6.4】 信源离散无记忆，输出符号 $X=\{0,1\}$，信道输出符号 $Y=\{0,1\}$，失真测度为 Hamming 失真测度

$$d(0,0)=d(1,1)=0$$
$$d(0,1)=d(1,0)=1$$

对信源做二次扩展 $\boldsymbol{X}=X_1X_2$，经离散无记忆有扰信道传输，输出符号 $\boldsymbol{Y}=Y_1Y_2$，计算扩展后的失真测度矩阵。

根据式（6-2），可计算出

$$d^{(2)}(00,00)=\frac{1}{2}[d(0,0)+d(0,0)]=0$$
$$d^{(2)}(01,00)=\frac{1}{2}[d(0,0)+d(1,0)]=\frac{1}{2}$$

类似地可计算出其他序列的失真测度，写成矩阵形式为

$$[d^{(2)}]=\begin{array}{c}\\ 00\\ 01\\ 10\\ 11\end{array}\begin{array}{c}\begin{array}{cccc}00 & 01 & 10 & 11\end{array}\\ \begin{bmatrix}0 & 0.5 & 0.5 & 1\\ 0.5 & 0 & 1 & 0.5\\ 0.5 & 1 & 0 & 0.5\\ 1 & 0.5 & 0.5 & 0\end{bmatrix}\end{array} \tag{6-3}$$

2. 平均失真

离散信源$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_I \\ q(x_1) & q(x_2) & \cdots & q(x_I) \end{bmatrix}$，经有扰信道传输，信道输出符号为 $Y=\{y_1, y_2, \cdots, y_J\}$，平均失真即对 d_{ij}（$i=1, 2, \cdots, I$; $j=1, 2, \cdots, J$）求统计平均值，记为

$$\overline{D}=\sum_{i=1}^{I}\sum_{j=1}^{J}p(x_iy_j)\,d_{ij}=\sum_{i=1}^{I}\sum_{j=1}^{J}q(x_i)p(y_j|x_i)\,d_{ij} \tag{6-4}$$

平均失真 $\overline{D}$ 是对在给定信源分布 $q(x)$条件下，通过有扰信道传输而引起失真的统计平均度量。

【例 6.5】 等概信源$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$，通过信道转移概率矩阵为$\boldsymbol{P}=\begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.25 & 0.5 & 0.25 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$的信道传输，失真测度为平方误差失真测度，求平均失真。

由式（6-4）可计算得

$$\begin{aligned}\overline{D}&=\sum_{i=1}^{3}\sum_{j=1}^{3}q(x_i)p(y_j|x_i)\,d_{ij}\\&=\frac{1}{3}\sum_{i=1}^{3}\sum_{j=1}^{3}p(y_j|x_i)\,d_{ij}\\&=\frac{1}{3}[1\times0.2+4\times0.2+1\times0.25+1\times0.25+4\times0.1+1\times0.1]=\frac{2}{3}\end{aligned}$$

对于矢量传输情况，若信道的输入、输出符号均为 N 长序列 $\boldsymbol{X}=X_1\cdots X_k\cdots X_N$, $\boldsymbol{Y}=Y_1\cdots Y_k\cdots Y_N$, $X_k\in\{x_1, x_2, \cdots, x_I\}$，$Y_k\in\{y_1, y_2, \cdots, y_J\}$，平均失真定义为

$$\overline{D^{(N)}}=\frac{1}{N}\sum_{k=1}^{N}\overline{D_k}=\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{I}\sum_{j=1}^{J}p(x_{ki}y_{kj})d(x_{ki}, y_{kj}) \tag{6-5}$$

式中，$\overline{D_k}$ 表示序列中第 k 个符号的平均失真。式（6-5）说明离散无记忆 N 次扩展信道的输入符号 $\boldsymbol{X}=X_1\cdots X_k\cdots X_N$ 和输出符号 $\boldsymbol{Y}=Y_1\cdots Y_k\cdots Y_N$ 之间的平均失真，等于单个符号 x_{ki} 与 y_{kj}之间失真统计值的总和。

若矢量信源是原离散无记忆信源的 N 次扩展，且矢量信道也是原离散无记忆信道的 N 次扩展，则每个 $\overline{D_k}$（$k=1, 2, \cdots, N$）对一维信源信道所取的均值相等，即 $\overline{D_1}=\cdots=\overline{D_k}=\cdots=\overline{D_N}=\overline{D}$，从而 $\overline{D^{(N)}}=\overline{D}$。

【例 6.6】 离散无记忆信源$\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 \\ 1/2 & 1/2 \end{bmatrix}$，离散无记忆信道$\boldsymbol{P}=\begin{bmatrix} 3/4 & 1/4 \\ 1/3 & 2/3 \end{bmatrix}$，失真测度为 Hamming 失真测度，由式（6-4）可计算得

$$\begin{aligned}\overline{D}&=\sum_{i}\sum_{j}q(x_i)p(y_j|x_i)\,d_{ij}\\&=\frac{1}{2}\left(0\times\frac{3}{4}+1\times\frac{1}{4}+1\times\frac{1}{3}+0\times\frac{2}{3}\right)=\frac{7}{24}\end{aligned} \tag{6-6}$$

对上述信源做二维扩展，可得扩展信源概率分布

$$\begin{bmatrix} X \\ q(X) \end{bmatrix} = \begin{bmatrix} x_0x_0 & x_0x_1 & x_1x_0 & x_1x_1 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} \tag{6-7}$$

扩展信道概率分布

$$\boldsymbol{P} = \begin{bmatrix} p(y_0y_0|x_0x_0) & p(y_0y_1|x_0x_0) & p(y_1y_0|x_0x_0) & p(y_1y_1|x_0x_0) \\ p(y_0y_0|x_0x_1) & p(y_0y_1|x_0x_1) & p(y_1y_0|x_0x_1) & p(y_1y_1|x_0x_1) \\ p(y_0y_0|x_1x_0) & p(y_0y_1|x_1x_0) & p(y_1y_0|x_1x_0) & p(y_1y_1|x_1x_0) \\ p(y_0y_0|x_1x_1) & p(y_0y_1|x_1x_1) & p(y_1y_0|x_1x_1) & p(y_1y_1|x_1x_1) \end{bmatrix}$$

对离散无记忆信道有 $p(\boldsymbol{y}|\boldsymbol{x}) = p(y_1|x_1)\,p(y_0|x_0)$，因此可算出

$$\boldsymbol{P} = [p(\boldsymbol{y}|\boldsymbol{x})] = \begin{bmatrix} 9/16 & 3/16 & 3/16 & 1/16 \\ 3/12 & 6/12 & 1/12 & 2/12 \\ 3/12 & 1/12 & 6/12 & 2/12 \\ 1/9 & 2/9 & 2/9 & 4/9 \end{bmatrix} \tag{6-8}$$

将扩展后的二维信源及二维信道视为“单符号”信源信道，利用式（6-3）、式（6-7）、式（6-8）的计算结果，根据式（6-4）可算出

$$\begin{aligned} \overline{D^{(2)}} &= \sum_{\boldsymbol{X}}\sum_{\boldsymbol{Y}} p(\boldsymbol{Y}|\boldsymbol{X})q(\boldsymbol{X})d(\boldsymbol{X},\boldsymbol{Y}) \\ &= \frac{1}{4}\left[0.5\times\left(\frac{3}{16}+\frac{3}{16}+\frac{3}{12}+\frac{2}{12}+\frac{3}{12}+\frac{2}{12}+\frac{2}{9}+\frac{2}{9}\right)+1\times\left(\frac{1}{16}+\frac{1}{12}+\frac{1}{12}+\frac{1}{9}\right)\right] = \frac{7}{24} \end{aligned} \tag{6-9}$$

或者根据式（6-5），利用式（6-6）的计算结果，也可算得

$$\overline{D^{(2)}} = \frac{1}{2}\sum_{k=1}^{2}\overline{D_k} = \frac{1}{2}\left(\frac{7}{24}+\frac{7}{24}\right) = \frac{7}{24} \tag{6-10}$$

式（6-9）和式（6-10）的计算结果是一致的，因此对离散无记忆 N 维扩展序列的讨论可归结为对一维的讨论。

6.2 信息率失真函数$R(D)$

6.2.1 率失真函数的定义

给定信源，即信源概率分布 $q(x)$ 一定，给定失真测度矩阵$[d]=[d_{ij}]$，寻找信道，记它的转移概率矩阵为 $\boldsymbol{P} = [p(y_j|x_i)]$，要求满足

$$\overline{D} = \sum_i\sum_j q(x_i)p(y_j|x_i)d_{ij} \le D \tag{6-11}$$

式中，D 是预先给定的失真度。式（6-11）称为保真度准则。

根据定理 2.2，当信源 $q(x)$一定时，平均互信息量 $I(X;Y)$是信道转移概率函数 $p(y|x)$的∪型凸函数，这意味着可以关于 $p(y|x)$对平均互信息量 $I(X;Y)$求得极小值，定义这个极小值为率失真函数 $R(D)$，即

$$R(D) \triangleq \min_{p(y|x)}\left\{I(X;Y):\ \overline{D} \le D\right\} \tag{6-12}$$

我们现在讨论的是，在允许一定失真前提下，对信源进行有失真编码，编码方法不同相当

于 $p(y|x)$不同。但这里的 $p(y|x)$并不是真正的信道，对于真正的信道，要改变其特性 $p(y|x)$代价太大，所以式（6-12）中的 $p(y|x)$只是试验信道，是我们假想的信道，它对应于不同的信源编码。因此式（6-12）的意义在于，选择 $p(y|x)$即选择某种编码方法在满足 $\overline{D} \leq D$ 的前提下，使 $I(X;Y)$ 达到最小值 $R(D)$，这就是满足平均失真 $\overline{D} \leq D$ 条件下的信源信息量可压缩的最低程度。

6.2.2 率失真函数的值域、定义域

1. $R(D)$的值域

$R(D)$的值域如图 6-1 所示。由平均互信息量 $I(X;Y)$的定义式（2-35），有

$$I(X;Y) = H(X) - H(X|Y)$$

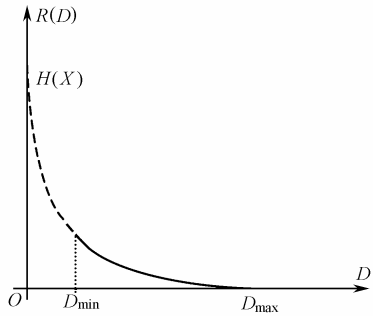


图 6-1 $R(D)$的值域

式中，$H(X|Y)$是信道干扰引起的疑义度，现在我们先避开信道干扰不谈，只讨论人为地对信源进行压缩编码带来的信息损失，所以上式变为 $I(X;Y) = H(X)$，这就说明在不允许任何失真的前提下，平均互信息量 $I(X;Y)$ 就等于信源熵 $H(X)$，若给定失真度 $D = 0$，就表示不允许任何失真，根据率失真函数的定义式（6-12）可知，在这种情况下，率失真函数 $R(D)$取得最大值 $H(X)$，即 $R_{\max}(D) = H(X)$。

由于 $R(D)$是 $I(X;Y)$在满足某些约束条件下的最小值，而由平均互信息量的性质 1 可知，$I(X;Y) \geqslant 0$，所以 $R(D) \geqslant 0$ 是非负函数，其最小值应为零，即 $R_{\min}(D) = 0$。

由上述讨论可知，率失真函数的值域为

$$0 \leqslant R(D) \leqslant H(X) \qquad (6\text{-}13)$$

2. $R(D)$的定义域

（1）D 的最小值 $D_{\min}$

由前面的讨论可知，在不允许失真的前提下，即 $D = 0$ 时，有 $R(D=0) = R_{\max}(D) = H(X)$。

D 能否达到其下限值零，与给定的单个符号的失真测度 $d(x_i, y_j)$有关，在给定的失真测度矩阵中，对每一个 x_i，找一个最小的 d_{ij}，然后对所有最小的 d_{ij}（$i = 1, 2, \cdots, I$）求统计平均值，就是 D 的最小值，即

$$D_{\min} \triangleq \sum_i q(x_i) \min_{y_j} d_{ij} \qquad (6\text{-}14)$$

从 $D_{\min}$ 的表达式可看出，只有在失真测度矩阵$[d]$中，每一行至少有一个零元素，才可能有 $D_{\min} = 0$，当 $D_{\min} = 0$ 时，表示不允许任何失真，此时 $R(D) = H(X)$（见图 6-1）。

（2）D 的最大值 $D_{\max}$

在允许一定失真的前提下，可对信源进行压缩编码，即传输的信息率可以小于信源熵，允许的失真越大，必需的信息率就越小。反过来说，必需的信息率越小，可承受的失真就越大，当 $R(D)$ 达到其最小值 $R_{\min}$（D）= 0 时，对应的失真最大，这种情况下 D 对应着 $R(D)$ 函数定义域的上界值 $D_{\max}$，如图 6-1 所示。

$$D_{\max} \triangleq \min\{D: R(D)=0\}$$
$$= \min\{D: I(X;Y)=0\} \tag{6-15}$$

当信源符号 x_i 与信宿符号 y_j 不相干时，有 $p(y_j|x_i) = \omega(y_j)$，此时，经过一次通信得不到任何信息，平均互信息量 $I(U;V)=0$，式（6-15）可改写为

$$D_{\max} = \min\{D: p(y_j|x_i) = \omega(y_j)\} \qquad i=1,2,\cdots,I;\ j=1,2,\cdots,J \tag{6-16}$$

式（6-16）表示，在所有满足 $p(y_j|x_i) = \omega(y_j)$的 D 中，D 的最小值即为 $R(D)$定义域的上限 $D_{\max}$。

但式（6-16）并没有给出计算 $D_{\max}$ 的具体方法，下面进一步求出计算 $D_{\max}$ 的显式。

计算 $R(D)$的前提是 $\overline{D} \leq D$，所以 D 的最大值也就是 $\overline{D}$ 的最大值。

由式（6-16）得

$$D_{\max} = \min_{p(y_j|x_i)=\omega(y_j)} \overline{D}$$
$$= \min_{p(y_j|x_i)=\omega(y_j)} \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_j|x_i)q(x_i)d_{ij}$$
$$= \min_{\omega(y_j)} \sum_{j=1}^{J} \omega(y_j)\left(\sum_{i=1}^{I} q(x_i)d_{ij}\right) \tag{6-17}$$

关于 $\omega(y_j)$ 求式（6-17）右边的极小值，可以这样来选择 $\omega(y_j)$ 的分布。

$$\begin{cases} \sum_{i=1}^{I} q(x_i)\,d_{ij} \text{ 取最小值时，选 } \omega(y_j)=1 \\ \sum_{i=1}^{I} q(x_i)\,d_{ij} \text{ 为其他值时，选 } \omega(y_j)=0 \end{cases} \qquad j=1,2,\cdots,J$$

这种分布可使等式（6-17）右边达到最小。

这样根据式（6-17）就有

$$D_{\max} = \min_{j} \sum_{i=1}^{I} q(x_i)d_{ij},\ j=1,2,\cdots,J \tag{6-18}$$

式（6-18）就是求 $D_{\max}$ 的表达式，在失真测度矩阵中，将第 j（j=1, 2, …, J）列的每个元素 d_{ij} 乘以相应的 $q(x_i)$，再把它们加起来，就得到 J 个和值，在这 J 个和值中，找一个最小值，就是 $D_{\max}$。

综上所述，$R(D)$的定义域为 $D_{\min} \leqslant D \leqslant D_{\max}$，式中 $D_{\min}$ 和 $D_{\max}$ 可分别由式（6-14）和式（6-18）求出。

【例 6.7】 一信源含有三个消息，概率分布为 $q_1=0.2, q_2=0.3, q_3=0.5$，失真测度矩阵为 $[d]=\begin{bmatrix} 4 & 2 & 1 \\ 0 & 3 & 2 \\ 2 & 0 & 1 \end{bmatrix}$，求 $D_{\min}$ 和 $D_{\max}$。

由式（6-14）可求出 $D_{\min} = 1\times0.2 + 0\times0.3 + 0\times0.5 = 0.2$；

由式（6-18）可求出 $D_{\max}=\min\{4\times0.2+2\times0.5,\ 2\times0.2+3\times0.3,\ 1\times0.2+2\times0.3+1\times0.5\}$

$=\min\{1.8,\ 1.3,\ 1.3\}=1.3$。

6.2.3 率失真函数的性质

率失真函数有如下几条性质。

1. $R(D)$是D的∪型凸函数

分别给定两个失真度 D_1 和 D_2（$D_{\min}\leqslant D_1, D_2\leqslant D_{\max}$），则下式成立

$$R(\alpha_1D_1+\alpha_2D_2)\alpha_1R(D_1)+\alpha_2R(D_2) \tag{6-19}$$

式中，$\alpha_1, \alpha_2\geqslant0$，且$\alpha_1+\alpha_2=1$，式（6-19）说明函数的均值大于等于均值的函数。

证明：

给定信源 $q(x)$，给定失真测度矩阵$[d]$，假设已找到两个实验信道，其信道转移概率矩阵分别为 $p_1(y_j|x_i)$和 $p_2(y_j|x_i)$，$i=1,2,\cdots,I$；$j=1,2,\cdots,J$，这两个信道所对应的平均互信息量分别为 $I_1(X;Y)$和 $I_2(X;Y)$。

在满足$\begin{cases}\overline{D_1}=\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}p_1(y_j|x_i)q(x_i)d_{ij}\leqslant D_1\\ \overline{D_2}=\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}p_2(y_j|x_i)q(x_i)d_{ij}\leqslant D_2\end{cases}$的条件下，使得平均互信息量满足

$$\begin{cases}I_1(X;Y)=R(D_1)\\ I_2(X;Y)=R(D_2)\end{cases} \tag{6-20}$$

现在构造一个新的信道，其信道转移概率为 $p(y_j|x_i)=\alpha_1p_1(y_j|x_i)+\alpha_2p_2(y_j|x_i)$，用新信道的转移概率 $p(y_j|x_i)$来计算平均失真，即

$$\begin{aligned}&\sum_{i=1}^{I}\sum_{j=1}^{J}p(y_j|x_i)q(x_i)d_{ij}\\ &=\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\alpha_1p_1(y_j|x_i)+\alpha_2p_2(y_j|x_i)\right]q(x_i)d_{ij}\\ &=\alpha_1\sum_{i=1}^{I}\sum_{j=1}^{J}p_1(y_j|x_i)q(x_i)d_{ij}+\alpha_2\sum_{i=1}^{I}\sum_{j=1}^{J}p_2(y_j|x_i)q(x_i)d_{ij}\\ &=\alpha_1\overline{D}_1+\alpha_2\overline{D}_2\xlongequal{\text{记为}}\overline{D}\end{aligned} \tag{6-21}$$

在率失真函数的定义 $R(D)\triangleq\min\limits_{p(y|x)}\{I(X;Y):\ \overline{D}\leqslant D\}$ 中，要求满足条件 $\overline{D}\leqslant D$，取式中的等号，即取 $\overline{D}=D$，则式（6-21）变为

$$\alpha_1D_1+\alpha_2D_2=D \tag{6-22}$$

由定理 2.2 知，当信源 $q(x)$给定时，平均互信息量 $I(X;Y)$ 是信道转移概率 $p(y|x)$的∪型凸函数，即下式成立

$$I(X;Y)\leqslant\alpha_1I_1(X;Y)+\alpha_2I_2(X;Y) \tag{6-23}$$

将式（6-20）代入式（6-23），有

$$I(X;Y)\leqslant\alpha_1R(D_1)+\alpha_2R(D_2) \tag{6-24}$$

又由率失真函数的定义，并利用式（6-22），有

$$R(\alpha_1 D_1 + \alpha_2 D_2) = R(D) \leqslant I(X;Y) \tag{6-25}$$

综合式（6-24）和式（6-25），有

$$R(\alpha_1 D_1 + \alpha_2 D_2) \leqslant \alpha_1 R(D_1) + \alpha_2 R(D_2)$$

证毕

2. $R(D)$是D的连续、单调、减函数

证明：

（1）$R(D)$的连续性可由 $I(X;Y) = I[q(x), p(y|x)]$是 $p(y|x)$的连续函数得出；

（2）$R(D)$的减函数性质好理解，因为允许的失真越大，则可选择的实验信道范围越大，对应的信息率可以越小；

（3）下面证明 $R(D)$是严格单调递减函数，即当 $D_1 < D_2$ 时，有 $R(D_1) > R(D_2)$。

设对于给定的失真度 D_1，D_2 和 $D_{\max}$（$0 < D_1 < D_2 < D_{\max}$），已找到对应的信道转移概率矩阵 $p_1(y|x)$，$p_2(y|x)$和 $p_0(y|x)$，使得对应的平均互信息量等于各自的率失真函数，即有

$$\begin{cases} R(D_1) = I[p_1(y|x)] \\ R(D_2) = I[p_2(y|x)] \\ R(D_{\max}) = I[p_0(y|x)] = 0 \end{cases}$$

令 $D_\theta = (1-\theta)D_1 + \theta D_{\max}$（$\theta$ 足够小，使 $D_1 < D_\theta < D_2$）

对应信道转移概率 $p_\theta(y|x) = (1-\theta)\, p_1(y|x) + \theta p_0(y|x)$

则有

$$\begin{aligned} R(D_\theta) &= \min_{p_\theta(v|u)} I[p_\theta(v|u)] \\ &\leqslant (1-\theta) I[p_1(y|x)] + \theta I[p_0(y|x)] \\ &= (1-\theta) I[p_1(y|x)] \\ &< I[p_1(y|x)] \\ &= R(D_1) \end{aligned}$$

可见，当 $D_1 < D_\theta$ 时，有

$$R(D_1) > R(D_\theta)$$

证毕

综上所述，可画出 $R(D)$-D 曲线，如图 6-2 所示。

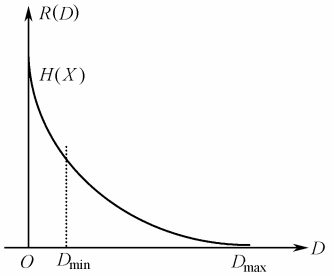


图 6-2　$R(D)$-D 曲线

3. 对于离散无记忆信源（DMS），$R^{(N)}(D) = N R(D)$

$R(D)$是每次传送一个符号时的率失真函数，$R^{(N)}(D)$ 是每次传送 N 个符号时的率失真函数。

证明：

（1）给定 $R^{(N)}(D)$，假设找到一个信道，信道输入为符号序列 $\boldsymbol{x} = x_1 x_2 \cdots x_N$，输出为符号序列 $\boldsymbol{y} = y_1 y_2 \cdots y_N$，其信道转移概率为 $p(\boldsymbol{y}|\boldsymbol{x})$，使得对应的平均互信息量 $I(\boldsymbol{X};\boldsymbol{Y}) = R^{(N)}(D)$，且平均失真测度 $\overline{D^{(N)}} \leq ND$，欲证 $R^{(N)}(D) \geqslant N R(D)$。

因为信源离散无记忆，由式（2-55）有

$$I(\boldsymbol{X};\boldsymbol{Y}) \geq \sum_{i=1}^{N} I(X_i;Y_i)$$
$$\geq \sum_{i=1}^{N} R(D_i)$$
$$= N\sum_{i=1}^{N}\frac{1}{N}R(D_i) \tag{6-26}$$

式中，D_i 是传送 $\boldsymbol{x}$ 中第 i 个符号的失真度，由于 $\boldsymbol{x}$ 中 N 个符号都在同一信道中传输，故有

$$D_1 = D_2 = \cdots = D_N = D$$

由于 $R(D)$是 D 的∪型凸函数，由式（6-26）得

$$I(\boldsymbol{X};\boldsymbol{Y}) \geq NR\left(\sum_{i=1}^{N}\frac{1}{N}D_i\right) \tag{6-27}$$

根据率失真函数的定义式（6-12），要求满足 $\overline{D_i} \leq D_i$，取式中等号，即取 $D_i = \overline{D_i} = E\{d(x_i,y_i)\}$，则式（6-27）中

$$\sum_{i=1}^{N}\frac{1}{N}D_i = \frac{1}{N}\sum_{i=1}^{N}E\{d(x_i,y_i)\}$$
$$= \frac{1}{N}E\left\{\sum_{i=1}^{N}d(x_i,y_i)\right\}$$
$$= \frac{1}{N}E\{d(\boldsymbol{x},\boldsymbol{y})\}$$
$$\leq \frac{1}{N}\cdot ND = D$$

上式中 $E\{d(\boldsymbol{x},\boldsymbol{y})\}$是信道传输 N 长序列时的总平均失真，将 $\sum_{i=1}^{N}\frac{1}{N}D_i \leq D$ 代入式（6-27），考虑到 $R(D)$是 D 的递减函数，得

$$I(\boldsymbol{X};\boldsymbol{Y}) \geqslant N R(D)$$

即

$$R^{(N)}(D) \geqslant N R(D) \tag{6-28}$$

（2）给定 $R(D)$，假设找到一个信道，其信道转移概率为 $p(y \mid x)$，使得对应的平均互信息量 $I(X;Y) = R(D)$，且平均失真测度 $E\{d(x,y)\} \leqslant D$，用信道 $p(y \mid x)$来传送 N 个符号，这 N 个符号互不相关（给定离散无记忆信源，构造一个离散无记忆信道），总平均失真

$$E\{d(\boldsymbol{x},\boldsymbol{y})\} = E\left\{\sum_{i=1}^{N}d(x_i,y_j)\right\} \leq ND$$

由于信源、信道离散无记忆，则根据定理 2.3 和定理 2.4 有

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=1}^{N} I(X_i;Y_i) = NR(D) \tag{6-29}$$

由于 $\boldsymbol{x}$ 中 N 个符号都在同一信道中传输，有 $I(X_1;\ Y_1) = I(X_2;\ Y_2) = \cdots = I(X_N;\ Y_N) = R(D)$。

又根据定义有

$$I(\boldsymbol{X};\boldsymbol{Y}) \geqslant R^{(N)}(D) \tag{6-30}$$

将式（6-30）代入式（6-29），得

$$NR(D) \geqslant R^{(N)}(D) \tag{6-31}$$

综合式（6-28）和式（6-31），得

$$R^{(N)}(D) = N\,R(D)$$ 证毕

因此，对于离散无记忆信源 DMS，我们只关心一维的情况。

6.3 率失真函数的计算

6.3.1 两种特殊情况下的求解

先通过以下两个例子来介绍 $R(D)$的计算方法。

【例 6.8】 信源含两个消息$\{x_1=0，x_2=1\}$，其概率分布为$\begin{bmatrix} X \\ p(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ \delta & 1-\delta \end{bmatrix}$，$\delta<0.5$，信道输出符号 $Y=\{y_1=0，y_2=1\}$，失真测度为汉明（Hamming）失真测度，求率失真函数 $R(D)$。

（1）根据式（6-14）和式（6-18）可求出 $R(D)$的定义域

$$D_{\min}=0\cdot\delta+0\cdot(1-\delta)=0$$

$$D_{\max}=\min\{1-\delta,\ \delta\}=\delta$$

（2）求 $R(D)$的值域

$$R(D_{\min}=0)=H(X)=-\delta\log\delta-(1-\delta)\log(1-\delta)=H_2(\delta)$$

$$R(D_{\max})=R(\delta)=0$$

（3）在 $0\leqslant D\leqslant\delta$的范围内，计算 $R(D)$

根据互信息量的表达式

$$I(X;Y)=H(X)-H(X\mid Y)$$

$$I(X;Y)=H_2(\delta)-H(X|Y) \qquad (6\text{-}32)$$

又根据 Fano 不等式有 $H(X\mid Y)\leqslant H_2(p_e)+p_e\log(k-1)$

此例中 $k=2$， $H(X\mid Y)\leqslant H_2(p_e)+p_e\log(2-1)$

将式（6-33）代入式（6-32）得

$$I(X;Y)\geqslant H_2(\delta)-H_2(p_e) \qquad (6\text{-}33)$$

式中，p_e为传输错误概率，又根据定义

$$R(D)\triangleq\min_{p(y|x)}\left\{I(X;Y):\ \overline{D}\leq D\right\}=H_2(\delta)-H_2(p_e)$$

下面找出 p_e与 D 的关系。

记信道转移概率矩阵 $\boldsymbol{P}=\begin{bmatrix} p(x_1|y_1) & p(x_1|y_2) \\ p(x_2|y_1) & p(x_2|y_2) \end{bmatrix}$

信道输入 x_i，输出 y_j，i，$j=1$，2，若 $i\neq j$，则认定传输出错，故

$$p_e=q(x_1)\,p(y_2\mid x_1)+q(x_2)\,p(y_1\mid x_2)$$

取 $D=\overline{D}=q(x_1)\left[p(y_1\mid x_1)d_{11}+p(y_2\mid x_1)d_{12}\right]+q(x_2)\left[p(y_1\mid x_2)d_{21}+p(y_2\mid x_2)d_{22}\right]$

$$=q(x_1)\left[p(y_1\mid x_1)\cdot 0+p(y_2\mid x_1)\cdot 1\right]+q(x_2)\left[p(y_1\mid x_2)\cdot 1+p(y_2\mid x_2)\cdot 0\right]$$

$$=q(x_1)\,p(y_2\mid x_1)+q(x_2)\,p(y_1\mid x_2)$$

$$=p_e$$

从而有 $R(D)=H_2(\delta)-H_2(p_e)=H_2(\delta)-H_2(D)$

（4）上面按照定义 $R(D)=\min_{p(v|u)}\left\{I(U;V):\overline{D}\leq D\right\}$求出了 $R(D)$，下面的问题是要真正找到

一个信道转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix} p(x_1|y_1) & p(x_1|y_2) \\ p(x_2|y_1) & p(x_2|y_2) \end{bmatrix}$ 的信道，使 $H(X \mid Y)= H_2(D)$，从而 $R(D)= H_2(\delta)-H_2(D)$，且 $\boldsymbol{P}$ 中的每一个元素 $p(y_j \mid x_i)$ 都满足

$$p(y_j \mid x_i) \leqslant 0 \qquad i,j=1,2$$

$$\sum_{j=1}^{2} p(y_j|x_i)=1$$

最方便的方法就是根据失真测度矩阵[d]的对称性，假设一个反向信道($Y \to X$)，反向信道的转移概率矩阵为

$$\left[\phi(x|y)\right]=\begin{bmatrix} \phi(x_1|y_1) & \phi(x_1|y_2) \\ \phi(x_2|y_1) & \phi(x_2|y_2) \end{bmatrix}=\begin{bmatrix} 1-D & D \\ D & 1-D \end{bmatrix}$$

由假设的反向信道计算平均失真，得

$$\begin{aligned} \overline{D} &= \sum_{i=1}^{2}\sum_{j=1}^{2}\phi(x_i|y_j)\omega(y_j)d_{ij} \\ &= \omega(y_1)[\,(1-D)\times 0 + D\times 1\,] + \omega(y_2)[\,D\times 1 + (1-D)\times 0\,] \\ &= [\,\omega(y_1)+\omega(y_2)\,]D \\ &= D \end{aligned}$$

由上式知 $\overline{D}=D$，满足失真条件 $\overline{D} \leqslant D$。

由假设的反向信道计算条件熵：

$$\begin{aligned} H(X|Y) &= -\sum_{i=1}^{2}\sum_{j=1}^{2}\omega(y_j)\phi(x_i|y_j)\log\phi(x_i|y_j) \\ &= -\omega(y_1)\,[(1-D)\log(1-D)+D\log D\,] - \omega(y_2)\,[D\log D+(1-D)\log(1-D)] \\ &= -(1-D)\log(1-D) - D\log D \\ &= H_2(D) \end{aligned}$$

则平均互信息量 $I(X;Y)=H(X)-H(X \mid Y)=H_2(\delta)-H_2(D)$，可见，假设的反向信道[$\phi(x \mid y)$]确实在满足 $\overline{D}=D$ 的条件下，使 $I(X;Y)=H_2(\delta)-H_2(D)$。

从而有 $$R(D)=\begin{cases} H_2(\delta)-H_2(D) & 0 \leqslant D \leqslant \delta \\ 0 & D > \delta \end{cases}$$

（5）要找出正向信道，可由 $q(u_i)=\sum_{j=1}^{2}\phi(u_i|v_j)\omega(v_j),\ i=1,2$，反解出 $\omega(y_j)$，$j=1,2$，再计算出 $p(y_j|x_i)=\dfrac{\phi(x_i|y_j)\omega(y_j)}{q(x_i)}$。

① $\delta=(1-D)\omega(y_1)+D\omega(y_2)$

$1-\delta=D\omega(y_1)+(1-D)\omega(y_2)$

由上面的方程组解出 $\omega(y_1)=\dfrac{\delta-D}{1-2D}$，$\omega(y_2)=\dfrac{1-\delta-D}{1-2D}$。

② 再算出

$$p(y_1|x_1)=\frac{\phi(x_1|y_1)\omega(y_1)}{q(x_1)}=\frac{(1-D)\dfrac{\delta-D}{1-2D}}{\delta}=\frac{(1-D)(\delta-D)}{\delta(1-2D)}$$

$$p(y_2|x_1)=\frac{\phi(x_1|y_2)\omega(y_2)}{q(x_1)}=\frac{D\cdot\frac{1-\delta-D}{1-2D}}{\delta}=\frac{D(1-\delta-D)}{\delta(1-2D)}$$

$$p(y_1|x_2)=\frac{\phi(x_2|y_1)\omega(y_1)}{q(x_2)}=\frac{D\cdot\frac{\delta-D}{1-2D}}{1-\delta}=\frac{D(\delta-D)}{(1-\delta)(1-2D)}$$

$$p(y_2|x_2)=\frac{\phi(x_2|y_2)\omega(y_2)}{q(x_2)}=\frac{(1-D)\cdot\frac{1-\delta-D}{1-2D}}{1-\delta}=\frac{(1-D)(1-\delta-D)}{(1-\delta)(1-2D)}$$

【例 6.9】 信源分布$\begin{bmatrix}X\\q(X)\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3\\1/3 & 1/3 & 1/3\end{bmatrix}$，失真测度矩阵为$[d]=\begin{matrix}x_1\\x_2\\x_3\end{matrix}\begin{matrix}y_1 & y_2\end{matrix}\begin{bmatrix}1 & 2\\1 & 1\\2 & 1\end{bmatrix}$，计算率失真函数$R(D)$。

（1）根据式（6-14）和式（6-18）可求出 $R(D)$的定义域

$$D_{\min}=\frac{1}{3}(1+1+1)=1,\ D_{\max}=\min\left\{\frac{1}{3}\times 4,\ \frac{1}{3}\times 4\right\}=\frac{4}{3}$$

（2）计算 $R(D)$

在信源等概的情况下，若失真测度矩阵$[d]$具有对称性，则在 $p(y\mid x)$具有相同对称性时，求出的 $I(X;Y)$ 就等于率失真函数 $R(D)$。

在此例中根据$[d]$的对称性，可假设信道转移概率矩阵$\boldsymbol{P}=\begin{bmatrix}1-\alpha & \alpha\\1/2 & 1/2\\\alpha & 1-\alpha\end{bmatrix}$，式中，$\alpha$ 为待定常数。

由假设的信道转移概率计算信息量

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=\sum_{j=1}^{2}\omega(y_j)\log\frac{1}{\omega(y_j)}-\sum_{i=1}^{3}\sum_{j=1}^{2}p(y_j|x_i)q(x_i)\log\frac{1}{p(y_j|x_i)}\end{aligned}\qquad(6\text{-}34)$$

先算出

$$\begin{cases}\omega(y_1)=\sum_{i=1}^{3}p(y_1|x_i)q(x_i)=\frac{1}{3}\left(1-\alpha+\frac{1}{2}+\alpha\right)=\frac{1}{2}\\\omega(y_2)=1-\omega(y_1)=\frac{1}{2}\end{cases}\qquad(6\text{-}35)$$

将式（6-35）代入式（6-34）得

$$\begin{aligned}I(X;Y)&=\log 2+\frac{1}{3}\sum_{i=1}^{3}\sum_{j=1}^{2}p(y_j|x_i)\log p(y_j|x_i)\\&=\log 2+\frac{1}{3}\times 2\times\left[(1-\alpha)\log(1-\alpha)+\alpha\log\alpha+\frac{1}{2}\log\frac{1}{2}\right]\\&=\frac{2}{3}[\log 2-H_2(\alpha)]\end{aligned}$$

即

$$I(X;Y)=\frac{2}{3}[\log 2-H_2(\alpha)] \tag{6-36}$$

由假设的信道转移概率计算平均失真，得

$$\overline{D}=\sum_{i=1}^{3}\sum_{j=1}^{2}p(y_j|x_i)q(x_i)d_{ij}=\frac{1}{3}\times 2\times\left[(1-\alpha)+2\alpha+\frac{1}{2}\right]=1+\frac{2}{3}\alpha \tag{6-37}$$

因为 $\overline{D}\leq D$，由式（6-37）得 $1+\frac{2}{3}\alpha\leq D$

即 $$\alpha\leq \frac{3}{2}(D-1)$$

考虑到 $D_{\max}=\frac{4}{3}$，则 $$\alpha\leq \frac{3}{2}\left(\frac{4}{3}-1\right)=\frac{1}{2}$$

如图 6-3 所示，在 $0<\alpha<\frac{1}{2}$ 的范围内，$H_2(\alpha)$是单调递增函数，有 $H_2(\alpha)\leq H_2\left[\frac{3}{2}(D-1)\right]$

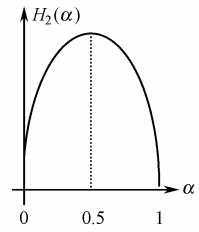


图 6-3　$H_2(\alpha)$—α曲线

根据式（6-36），得

$$I(X;Y)\geq \frac{2}{3}\left\{\log 2-H_2\left[\frac{3}{2}(D-1)\right]\right\}$$

从而有

$$R(D)=\begin{cases}\dfrac{2}{3}\left\{\log 2-H_2\left[\dfrac{3}{2}(D-1)\right]\right\} & 1\leq D\leq \dfrac{4}{3}\\ 0 & D<1,\ D>\dfrac{4}{3}\end{cases}$$

6.3.2　*R*(*D*)的参数表示法

上两例都是利用失真矩阵[*d*]的对称性算得 *R*(*D*)，对于一般情况，问题归结为选择适当的实验信道，使平均互信息量 $I(X;Y)=H(Y)-H(Y|X)$ 取极小值。

实验信道的转移概率用 $p(y_j|x_i)$ 表示，$p(y_j|x_i)$ 须满足约束条件：

$$\begin{cases}p(y_j|x_i)\geq 0\\ \sum_{j=1}^{J}p(y_j|x_i)=1 & i=1,2,\cdots,I\\ & j=1,2,\cdots,J\\ \sum_{i=1}^{I}\sum_{j=1}^{J}p(y_j|x_i)q(x_i)d_{ij}\leq D\end{cases} \tag{6-38}$$

可以先不考虑式（6-38）中的第一个约束条件，用拉格朗日乘因子法求解，不过解通常只能用参数形式来表达。

构造函数：

$$\Phi \triangleq I(X;Y)-sD-\sum_{i=1}^{I}\mu_i\sum_{j=1}^{J}p(y_j|x_i)$$

$$= -\omega(y_j)\log\omega(y_j) + \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_j|u_i)q(u_i)\log p(y_j|u_i) -$$

$$s\sum_i\sum_j p(y_j|x_i)\,q(x_i)d_{IJ} - \sum_i \mu_i \sum_j p(y_j|x_i)$$

式中，s,μ_i 为拉格朗日乘因子。

令
$$\frac{\partial \Phi}{\partial p(y_j|x_i)} = 0$$

为计算方便，对数取 e 为底。注意到

$$\omega(y_j) = \sum_i p(y_j|x_i)q(x_i) \quad \Rightarrow \quad \frac{\partial \omega(y_j)}{\partial p(y_j|x_i)} = q(x_i)$$

$$\frac{\partial \Phi}{\partial p(y_j|x_i)} = -q(x_i)\ln\omega(y_j) - q(x_i) + q(x_i)\ln p(y_j|x_i) + q(x_i) - sq(x_i)d_{ij} - \mu_i$$

$$= q(x_i)\ln\frac{p(y_j|x_i)}{\omega(y_j)} - sq(x_i)d_{ij} - \mu_i = 0 \tag{6-39}$$

由式（6-39）得

$$p(y_j|x_i) = \omega(y_j)\exp\left\{\frac{sq(x_i)d_{ij} + \mu_i}{q(x_i)}\right\}$$

$$= \omega(y_j)\exp\left\{\frac{\mu_i}{q(x_i)}\right\}\exp\{sd_{ij}\}$$

$$= \omega(y_j)\lambda_i \mathrm{e}^{sd_{ij}} \tag{6-40}$$

式中，$\lambda_i = \exp\left\{\dfrac{\mu_i}{q(x_i)}\right\}$ 为待定常数。

将式（6-40）代入式（6-38）中的第二个约束条件（归一化条件），得

$$1 = \sum_{j=1}^{J} p(y_j|x_i) = \lambda_i \sum_{j=1}^{J} \omega(y_j)\mathrm{e}^{sd_{ij}}$$

$$\lambda_i = \frac{1}{\sum_{j=1}^{J}\omega(y_j)\mathrm{e}^{sd_{ij}}} \tag{6-41}$$

根据式（6-40）有

$$\omega(y_j) = \sum_{i=1}^{I} p(y_j|x_i)q(x_i) = \omega(y_j)\sum_{i=1}^{I}\lambda_i \mathrm{e}^{sd_{ij}}q(x_i)$$

即

$$1 = \sum_i \lambda_i \mathrm{e}^{sd_{ij}}q(x_i) \qquad j = 1,2,\cdots,J \tag{6-42}$$

下面求 $R(D)$ 的参数表达式，取 $D = \overline{D}$ 。

$$D = \overline{D} = \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_j|x_i)\,q(x_i)\,d_{ij}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J}\omega(y_j)q(x_i)\lambda_i d_{ij}\mathrm{e}^{sd_{ij}} \tag{6-43}$$

$$
\begin{aligned}
R(D) &= \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_j|x_i)q(x_i)\log\frac{p(y_j|x_i)}{\omega(y_j)} \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} p(y_j|x_i)q(x_i)(sd_{ij}+\log\lambda_i) \\
&= s\overline{D} + \sum_{j=1}^{J} p(y_j|x_i)\sum_{i=1}^{I} q(x_i)\log\lambda_i \\
&= sD + \sum_{i=1}^{I} q(x_i)\log\lambda_i
\end{aligned}
$$

即

$$
R(D) = sD + \sum_{i=1}^{I} q(x_i)\log\lambda_i \tag{6-44}
$$

可以证明，由式（6-44）决定的 $R(D)$满足 $\frac{\mathrm{d}R(D)}{\mathrm{d}D} = s$ 。

证明：

$$
\begin{aligned}
\frac{\mathrm{d}R(D)}{\mathrm{d}D} &= \frac{\partial R}{\partial D} + \frac{\partial R}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}D} + \sum_{i=1}^{I}\frac{\partial R}{\partial \lambda_I}\frac{\mathrm{d}\lambda_i}{\mathrm{d}D} \\
&= s + D\frac{\mathrm{d}s}{\mathrm{d}D} + \sum_{i=1}^{I}\frac{q(x_i)}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}D} \\
&= s + \left[D + \sum_{i=1}^{I}\frac{q(x_i)}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}\right]\frac{\mathrm{d}s}{\mathrm{d}D}
\end{aligned}
$$

下面证明方括弧中的 $D + \sum_{i=1}^{I}\frac{q(x_i)}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}$ 等于零，对式（6-42）两边关于 s 求微分，得

$$
\sum_{i=1}^{I} q(x_i)\left[\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}\mathrm{e}^{sd_{ij}} + \lambda_i \mathrm{e}^{sd_{ij}} d_{ij}\right] = 0 \tag{6-45}
$$

对式（6-45）两边乘以 $\omega(y_j)$，并对 j 求和，得

$$
\sum_{j=1}^{J}\omega(y_j)\sum_{i=1}^{I} q(x_i)\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}\mathrm{e}^{sd_{ij}} + \sum_{j=1}^{J}\omega(y_j)\sum_{i=1}^{I}\lambda_i q(x_i) d_{ij}\mathrm{e}^{sd_{ij}} = 0 \tag{6-46}
$$

将式（6-43）代入式（6-46），得

$$
\sum_{i=1}^{I} q(x_i)\frac{\mathrm{d}\lambda_i}{\mathrm{d}s}\sum_{j=1}^{J}\omega(y_j)\,\mathrm{e}^{sd_{ij}} + D = 0
$$

将式（6-41）代入上式，就有 $\sum_{i=1}^{I}\frac{q(x_i)}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}s} + D = 0$，这就证得了 $\frac{\mathrm{d}R(D)}{\mathrm{d}D} = s$，见图 6-4。

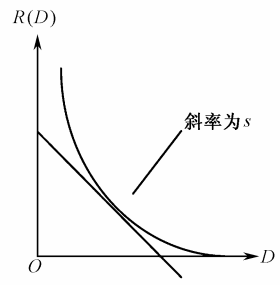


图 6-4 $\frac{\mathrm{d}R(D)}{\mathrm{d}D} = s$ 曲线

上面已求出了 $R(D)$的参数表达式，用参数法求 $R(D)$，可按下述步骤进行。

（1）由式（6-42）$1 = \sum_i \lambda_i \mathrm{e}^{sd_{ij}} q(x_i)$（$j = 1, 2, \cdots, J$）求 λ_i；

（2）由式（6-41）$\sum_{j=1}^{J}\omega(y_j)\mathrm{e}^{sd_{ij}} = \frac{1}{\lambda_i}$ 求 $\omega(v_j)$；

（3）由式（6-40） $p(y_j|x_i)=\omega(y_j)\lambda_i e^{sd_{ij}}$ 求 $p(y_j|x_i)$；

（4）由式（6-43） $D=\sum_{i=1}^{I}\sum_{j=1}^{J}\omega(y_j)q(x_i)\lambda_i d_{ij}e^{sd_{ij}}$ 求 D；

（5）由式（6-44） $R(D)=sD+\sum_{i=1}^{I}q(x_i)\log\lambda_i$ 求 $R(D)$。

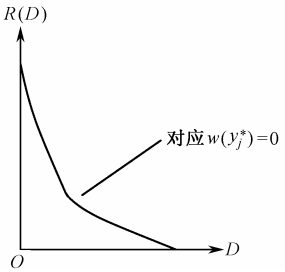


图 6-5　$\omega(y^*_j)<0$ 的情况

用此法求解，有时候会出现$\omega(y_j)$<0 的情况（对应 $p(y_j|x_i)<0$），这是因为用拉格朗日乘因子法求解时，并没有考虑式（6-38）中的第一个约束条件 $p(y_j|x_i)\geqslant 0$。碰到这种情况，就要令某一 $\omega(y^*_j)=0$，重复刚才的求解过程，这种情况下求得的 $R(D)$是一条折线，折点对应$\omega(y^*_j)=0$，如图 6-5 所示。

下面举一例说明如何用参数法求解。

【例 6.10】　仍考虑例 6.8 的输入概率分布 $q(x_1)=\delta$，$q(x_2)=1-\delta$，$\delta<0.5$，失真测度矩阵 $[d]=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$，用参数法求 $R(D)$。

解：

（1）根据式（6-42），即 $\sum_{i=1}^{2}\lambda_i e^{sd_{ij}}q(x_i)=1 \quad (j=1,2)$ 得

$$\begin{cases}\delta\cdot\lambda_1 e^{0}+(1-\delta)\lambda_2 e^{s}=1\\ \delta\cdot\lambda_1 e^{s}+(1-\delta)\lambda_2 e^{0}=1\end{cases}\Rightarrow\begin{cases}\delta\cdot\lambda_1+(1-\delta)e^{s}\lambda_2=1\\ \delta\cdot e^{s}\lambda_1+(1-\delta)\lambda_2=1\end{cases}$$

解出
$$\lambda_1=\frac{\begin{vmatrix}1 & (1-\delta)e^{s}\\ 1 & 1-\delta\end{vmatrix}}{\begin{vmatrix}\delta & (1-\delta)e^{s}\\ \delta\cdot e^{s} & 1-\delta\end{vmatrix}}=\frac{(1-e^{s})(1-\delta)}{\delta(1-\delta)(1-e^{2s})}=\frac{1}{\delta(1+e^{s})}$$

$$\lambda_2=\frac{\begin{vmatrix}\delta & 1\\ \delta\cdot e^{s} & 1\end{vmatrix}}{\begin{vmatrix}\delta & (1-\delta)e^{s}\\ \delta\cdot e^{s} & 1-\delta\end{vmatrix}}=\frac{\delta(1-e^{s})}{\delta(1-\delta)(1-e^{2s})}=\frac{1}{(1-\delta)(1+e^{s})}$$

（2）由式（6-41），即 $\lambda_i=\frac{1}{\sum_{j=1}^{2}\omega(y_j)e^{sd_{ij}}}$（$i=1,2$），得

$$\begin{cases}\delta(1+e^{s})=\omega(v_1)e^{0}+\omega(v_2)e^{s}\\ (1-\delta)(1+e^{s})=\omega(v_1)e^{s}+\omega(v_2)e^{0}\end{cases}\Rightarrow\begin{cases}\omega(v_1)+e^{s}\omega(v_2)=\delta(1+e^{s})\\ e^{s}\omega(v_1)+\omega(v_2)=(1-\delta)(1+e^{s})\end{cases}$$

解出
$$\omega(y_1)=\frac{\begin{vmatrix}\delta(1+e^{s}) & e^{s}\\ (1-\delta)(1+e^{s}) & 1\end{vmatrix}}{\begin{vmatrix}1 & e^{s}\\ e^{s} & 1\end{vmatrix}}=\frac{(1+e^{s})(\delta-e^{s}+\delta\cdot e^{s})}{1-e^{2s}}=\frac{\delta-e^{s}+\delta\cdot e^{s}}{1-e^{s}}$$

$$\omega(y_2)=\frac{\begin{vmatrix}1 & \delta(1+\mathrm{e}^s)\\ \mathrm{e}^s & (1-\delta)(1+\mathrm{e}^s)\end{vmatrix}}{\begin{vmatrix}1 & \mathrm{e}^s\\ \mathrm{e}^s & 1\end{vmatrix}}=\frac{(1+\mathrm{e}^s)(1-\delta-\delta\cdot\mathrm{e}^s)}{1-\mathrm{e}^{2s}}=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}$$

（3）由式（6-40），即 $p(y_j|x_i)=\omega(y_j)\lambda_i\mathrm{e}^{sd_{ij}}\ (i,j=1,2)$，得

$$p(y_1|x_1)=\frac{\delta-\mathrm{e}^s-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\frac{1}{\delta(1+\mathrm{e}^s)}\mathrm{e}^{s\cdot 0}=\frac{\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{\delta(1-\mathrm{e}^{2s})}$$

$$p(y_2|x_1)=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\frac{1}{\delta(1+\mathrm{e}^s)}\mathrm{e}^s=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{\delta(1-\mathrm{e}^{2s})}\mathrm{e}^s$$

$$p(y_1|x_2)=\frac{\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\frac{1}{(1-\delta)(1+\mathrm{e}^s)}\mathrm{e}^s=\frac{\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{(1-\delta)(1-\mathrm{e}^{2s})}\mathrm{e}^s$$

$$p(y_2|x_2)=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\frac{1}{(1-\delta)(1+\mathrm{e}^s)}\mathrm{e}^{s\cdot 0}=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{(1-\delta)(1-\mathrm{e}^{2s})}$$

（4）由式（6-43），即 $D=\sum_{i=1}^{2}\sum_{j=1}^{2}\omega(y_j)q(x_i)\lambda_i d_{ij}\mathrm{e}^{sd_{ij}}$，得

$$\begin{aligned}D&=\omega(y_2)q(x_1)\lambda_1 d_{12}\mathrm{e}^{sd_{12}}+\omega(y_1)q(x_2)\lambda_2 d_{21}\mathrm{e}^{sd_{21}}\\&=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\cdot\delta\cdot\frac{1}{\delta(1+\mathrm{e}^s)}\mathrm{e}^s+\frac{\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^s}\cdot(1-\delta)\cdot\frac{\mathrm{e}^s}{(1-\delta)(1+\mathrm{e}^s)}\\&=\frac{1-\delta-\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^{2s}}\mathrm{e}^s+\frac{\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^{2s}}\mathrm{e}^s\\&=\frac{1-\delta-\delta\cdot\mathrm{e}^s+\delta-\mathrm{e}^s+\delta\cdot\mathrm{e}^s}{1-\mathrm{e}^{2s}}\mathrm{e}^s\\&=\frac{1-\mathrm{e}^s}{1-\mathrm{e}^{2s}}\mathrm{e}^s=\frac{\mathrm{e}^s}{1+\mathrm{e}^s}\end{aligned}$$

则有

$$s=\ln\frac{D}{1-D}$$

（5）由式（6-44），即 $R(D)=sD+\sum_{i=1}^{2}q(x_i)\log\lambda_i$，得

$$\begin{aligned}R(D)&=D\ln\frac{D}{1-D}+\delta\ln\frac{1}{\delta(1+\mathrm{e}^s)}+(1-\delta)\ln\frac{1}{(1-\delta)(1+\mathrm{e}^s)}\\&=D\ln\frac{D}{1-D}-\delta\ln\delta-\delta\ln(1+\mathrm{e}^s)-(1-\delta)\ln(1-\delta)-(1-\delta)\ln(1+\mathrm{e}^s)\\&=D\ln\frac{D}{1-D}+H_2(\delta)-\ln\left(1+\frac{D}{1-D}\right)\\&=D\ln D+(1-D)\ln(1-D)+H_2(\delta)\\&=H_2(\delta)-H_2(D)\end{aligned}$$

与例 6.8 求得的结果一致。

6.4 率失真信源编码定理

对于离散无记忆平稳信源，给定允许失真 D，当信息传输率 $R > R(D)$时，只要信源序列 $\boldsymbol{x} = x_1 x_2 \cdots x_N$ 足够长，一定存在一种编码方法（对应一种试验信道），使平均失真满足 $\overline{D} \leq D + \varepsilon$，$\varepsilon$是任意小的正数。反之，若信息传输率 $R < R(D)$，则无论如何采用何种编码方法，必有 $\overline{D} \geq D$。

从上述定理的叙述可看出，在不考虑信道干扰引起失真的前提下，为提高传输效率，对给定信源进行压缩编码，势必引起信源失真，给定所能承受的失真 D，则在满足 $\overline{D} \leq D + \varepsilon$ 的条件下，$R(D)$是信息传输率可压缩的下界。

定理包括两部分，$R > R(D)$的情况称为保真度准则下的率失真编码定理，$R < R(D)$的情况称为逆定理。

上述定理谓之**香农第三定理**，只是告知码的存在性，定理本身并没有给出具体的编码方法。实际上，后人只是循着该定理提供的思路去寻找各种可行的编码方法，目前尚无合适的系统编码方法使传输率逼近香农第三定理给出的下界 $R(D)$。

本 章 小 结

本章介绍了在允许一定失真的条件下，对给定信源进行压缩编码，信息传输率可压缩的最低下界称为率失真函数 $R(D)$。

重点讨论了以下几个问题。

（1）失真测度。

可以理解为误码造成的损失，实际上发送同一符号而错成不同符号所造成的损失是不同的。

（2）平均失真 $\overline{D}$。

对给定信源 $q(x)$ 进行压缩编码，不同的编码方法对应不同的试验信道，可用信道转移概率 $p(y \mid x)$ 来描述该试验信道，用概率分布 $p(x\,y) = q(x)\,p(y \mid x)$对给定的失真测度求统计平均值就得到平均失真 $\overline{D}$。

（3）如何求 $R(D)$的定义域和值域。

（4）求率失真函数 $R(D)$。

本章通过两个例子讨论了两种特殊情况下 $R(D)$ 的求法，在一般的离散无记忆情况下，可用书中介绍的参数法求解。当然一般离散信源的率失真函数的计算是相当困难和冗长烦琐的，往往要借助计算机进行，可利用迭代法求解，这在不少书中有介绍。

（5）香农第三定理——保真度准则下的率失真编码定理，这是信息理论的重要定理之一，由于证明繁杂，本书略去了其证明过程，希望读者搞清楚它的物理意义。

思考题与习题

6.1　当率失真函数 $R(D)$取什么值的时候，表示不允许有任何失真？

6.2　信源在不允许失真时，其信息率所能压缩到的极限值是什么？当允许信源存在一定的失真时，其信息率所能压缩到的极限值又是什么？

6.3　对于离散无记忆信源，率失真函数 $R(D)$的最大值是什么？信源概率 $q(x)$如何分布才能达到该最大值？此时对应的平均失真 $\overline{D}$ 是多少？

6.4　率失真函数 $R(D)$的最小值是什么？信源概率 $q(x)$如何分布才能达到该最小值？此时对应的平均失真 $\overline{D}$ 是多少？

6.5　给定信源分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 \\ 0.5 & 0.25 & 0.25 \end{bmatrix}$ 和失真测度矩阵 $[d]=\begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}$，求率失真函数的定义域和值域。

6.6　等概信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$，接收符号集 $Y=\{y_0, y_1, y_2, y_3\}$，失真测度矩阵为 $[d]=\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$，求 $R(D)$的定义域和值域。

6.7　在例 6.9 中，已经算出 $R(D)$的定义域为 $D_{\min}=1, D_{\max}=3/4$，问分别选择什么样的信道可使平均失真 $\overline{D}=D_{\min}=1$ 及 $\overline{D}=D_{\max}=3/4$？

6.8　已知离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}$ 及失真测度矩阵$[d]$，证明率失真函数在零点的值 $R(0)=H(X)$（为信源熵）的充要条件是，失真测度矩阵$[d]$中每一行至少有一个元素为零，且每一列至多只有一个元素为零。

6.9　离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$，失真测度为 Hamming 失真测度。

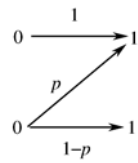
（1）求 $D_{\min}$ 和 $R(D_{\min})$，并写出对应试验信道的信道转移概率矩阵；

（2）求 $D_{\max}$ 和 $R(D_{\max})$，并写出对应试验信道的信道转移概率矩阵；

（3）若允许平均失真 $\overline{D}=\dfrac{1}{3}$，请问每个信源符号至少需要几个二进制符号来对其编码？

6.10　在例 6.8 中，当允许平均失真 $\overline{D}=0.5\delta$ 时，请问每个信源符号至少需要几个二进制符号来对其编码？

6.11　已知离散无记忆信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，二进制 Z 信道如题图 6-1 所示，失真测度为 Hamming 失真测度。



题图 6-1　二进制 Z 信道

（1）计算平均失真 $\overline{D}$；

（2）率失真函数 $R(D)$的最大值是什么？信源概率 $q(x)$如何分布才能达到该最大值？此时对应的平均失真 $\overline{D}$ 是多大？

（3）率失真函数 $R(D)$的最小值是什么？信源概率 $q(x)$如何分布才能达到该最小值？此时对应的平均失真 $\overline{D}$ 是多大？

（4）画出 $R(D)$~ D 曲线。

6.12 给定二元信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ 0.5 & 0.5 \end{bmatrix}$ 和失真测度矩阵 $[d]=\begin{bmatrix} 0 & \alpha \\ \alpha & 0 \end{bmatrix}$，求率失真函数 $R(D)$。

6.13 给定二元信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ \alpha & 1-\alpha \end{bmatrix}$，$\alpha<0.5$，以及失真测度矩阵 $[d]=\begin{bmatrix} 0 & \alpha \\ \alpha & 0 \end{bmatrix}$，求率失真函数 $R(D)$。

6.14 给定二元信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 \\ 0.5 & 0.5 \end{bmatrix}$，以及失真测度矩阵 $[d]=\begin{bmatrix} 0 & \alpha \\ \beta & 0 \end{bmatrix}$，$\alpha>0$，$\beta>0$，求率失真函数 $R(D)$。

6.15 给定信源符号集 $X=\{0,1\}$，信道符号集 $Y=\{0,e,1\}$，信源等概分布 $q(0)=q(1)=\frac{1}{2}$，失真测度矩阵 $[d]=\begin{bmatrix} 0 & \alpha & \beta \\ \beta & \alpha & 0 \end{bmatrix}$，式中 $\alpha<\frac{1}{2}$。

（1）求率失真函数 $R(D)$;

（2）当 $\alpha\geq\frac{1}{2},\beta=1$ 时，证明 $R(D)=\log 2-H_2(D)$。

6.16 给定二进制信道，信源分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，$p<0.5$，失真测度矩阵 $[d]=\begin{bmatrix} \beta & \alpha \\ \alpha & \beta \end{bmatrix}$，式中$\beta>\alpha$，求率失真函数 $R(D)$。

6.17 给定信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} 0 & 1 & 2 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$，失真测度为 Hamming 失真测度，求率失真函数 $R(D)$。

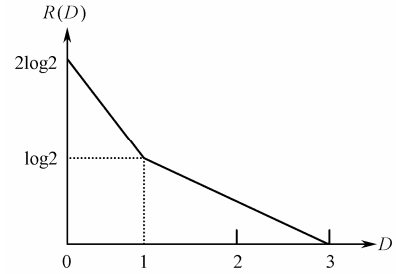
6.18 给定信源 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & x_3 \\ 2/5 & 1/5 & 2/5 \end{bmatrix}$，失真测度函数定义为 $d_{ij}=\begin{cases} 0 & i=j \\ 1 & i\neq j \end{cases}$（$i,j=1,2,3$），求率失真函数 $R(D)$;

6.19 信源等概分布 $\begin{bmatrix} X \\ q(X) \end{bmatrix}=\begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$，接收符号集 $Y=\{y_0,y_1,y_2,y_3\}$，失真测度矩阵为 $[d]=\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$。

（1）求 $R(D)$的定义域和值域;

（2）求 $R(D)$ 函数，并画出 $R(D)$～D 曲线（取 4～5 个点）。

6.20　信源 $X=\{x_1, x_2, x_3, x_4\}$，信宿 $Y=\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$，信源等概分布，失真测度矩阵为 $[d]=\begin{bmatrix} 0 & \infty & \infty & \infty & 1 & \infty & 3 \\ \infty & 0 & \infty & \infty & 1 & \infty & 3 \\ \infty & \infty & 0 & \infty & \infty & 1 & 3 \\ \infty & \infty & \infty & 0 & \infty & 1 & 3 \end{bmatrix}$，证明率失真函数 $R(D)$如题图 6-2 所示。



题图 6-2　率失真函数

6.21　信源符号集{0，1}，信宿符号集{0，1，2}，信源等概分布，失真测度矩阵 $[d]=\begin{bmatrix} 0 & \infty & 1 \\ \infty & 0 & 1 \end{bmatrix}$，求率失真函数 $R(D)$。

第 7 章
纠错编码代数基础

内容提要

编码理论是建立在码的代数结构基础上的。抽象代数为编码理论的发展提供了强有力的工具，其运算对象是数、多项式、矢量、矩阵、线性空间等。为便于后面纠错编码各章节的学习，本章简单介绍抽象代数中与编码直接相关的基础知识，包括线性空间和群、环、域的基本概念，群的陪集分解，有限域的性质等，为方便理解前面还介绍了整数及多项式的一些基本概念。

知识要点

群、环、域的基本概念，群的陪集分解，剩余类构成的有限域的性质及多项式 x^n-1 的因式分解。

教学建议

本章是进行后续各章学习的数学基础。建议学时数为 7 学时，其中基本概念一节可适当取舍，讲解中可以通过引入线性分组码或循环码的 1～2 个应用实例加深学生的理解和掌握。

7.1 基 本 概 念

7.1.1 整数

定理 7.1 （**带余除法定理**）设 a 为整数，d 为正整数，且 $a > d$，则存在唯一的整数 q，r 满足 $a = qd + r$，$0 \leqslant r < d$。

d 称为模，r 称为余数，r 可记为 $a \ [\mathrm{mod}\quad d\ ]$。

- **同余** 设 a，b 为整数，d 为正整数，若 $a = q_1 d + r$，$b = q_2 d + r$，$0 \leqslant r < d$ ，则称 a，b 关于模 d 同余，记为 $a \equiv b \ [\mathrm{mod}\quad d\ ]$。
- **剩余类** 模 d 运算余数相同的元素构成的集合为模 d 的剩余类，记为 $\bar{0}$，$\bar{1}$，…，$\overline{d-1}$，对应代表元常取 0，1，…，$d-1$。

剩余类之间也可定义加法和乘法运算：

$$\bar{a} + \bar{b} = \overline{a+b} \quad [\mathrm{mod}\quad d\ ]$$
$$\bar{a} \cdot \bar{b} = \overline{a \cdot b} \quad [\mathrm{mod}\quad d\ ]$$

【例 7.1】 $d = 7$，则

$$\bar{1} + \bar{2} = \overline{1+2} = \bar{3} \quad [\mathrm{mod}\quad 7\ ]$$
$$\bar{3} \times \bar{5} = \overline{3 \times 5} = \overline{15} = \bar{1} \quad [\mathrm{mod}\quad 7\ ]$$

模 d 的全体剩余类对模 d 加法和模 d 乘法满足封闭性，即假设 $D = \{\bar{0}, \bar{1}, \cdots, \overline{d-1}\}$，如果 $a, b \in D$，则必有 $(a + b)\ [\mathrm{mod}\quad d\ ] \in D$ 及 $(a \cdot b)\ [\mathrm{mod}\quad d\ ] \in D$。

为简化书写，常将模 d 的剩余类直接写为 0，1，…，$d-1$。

定理 7.2 （**算术基本定理**）任何正整数 a 均可表示成其素因数的幂之积：

$$a = p_1^{r_1} p_2^{r_2} \cdots p_n^{r_n}$$

$p_1, p_2, \cdots, p_n$ 为 a 的互不相同的素因数，r_i 为正整数。

若不考虑排列次序，这种分解是唯一的，如 $180 = 2^2 \times 3^2 \times 5^1$。

定理 7.3 （**最大公因数定理**）设 a，b 是不全为 0 的整数，则存在整数 p，q 使

$$pa + qb = (a, b)$$

式中，(a, b) 为 a，b 的最大公约数，当 a，b 互素时，$(a, b) = 1$，$pa + qb = 1$。

7.1.2 多项式

多项式的性质在很多方面类似于整数的性质。

系数取自集合 F 的多项式的表示形式为 $f(x) = f_n x^n + f_{n-1} x^{n-1} + \cdots + f_1 x + f_0$，$f_i \in F$。

- **首一多项式**：多项式的最高次数的系数为 1，即 $f_n = 1$。
- **多项式的阶**：多项式中系数不为 0 的 x 的最高次数，记为 $\partial^\circ f(x)$。
- **即约多项式**：阶大于 0 且在给定集合 F 上除了常数和常数与本身的乘积外，不能被其他多项式除尽的多项式。

【例 7.2】 $x^2 + 1$ 是阶为 2 的首一多项式，它在实数上是即约多项式，而在复数上不是即约多项式，因为在复数上可分解为两个因式（$x + i$）和（$x - i$）。

定理 7.4 给定任意两个多项式 $f(x)$，$p(x)$，$\partial^\circ f(x) > \partial^\circ p(x)$，一定存在唯一的多项式 $q(x)$

和 $r(x)$，使 $f(x)=q(x)\cdot p(x)+r(x)$，$0\leqslant\partial^{\circ}r(x)<\partial^{\circ}p(x)$。

$p(x)$称为模多项式，$r(x)$称为余式，$r(x)$记为 $f(x)[\bmod\ p(x)]$。

- **同余**：若 $a(x)=q_1(x)\cdot p(x)+r(x)$，$b(x)=q_2(x)\cdot p(x)+r(x)$，$0\leqslant\partial^{\circ}r(x)<\partial^{\circ}p(x)$，则 $a(x)\equiv b(x)[\bmod\ p(x)]$
- **剩余类**：模 $p(x)$运算余数相同的多项式集合，记为 $\overline{r(x)}$ 或 $r(x)$。

多项式的剩余类具有与整数同样的性质。

【例 7.3】 系数取自{0，1}的任意多项式以 $p(x)=x^3+x+1$ 为模，设所得余式为 $r(x)$，则有 $0\leqslant\partial^{\circ}r(x)<3$，令 $r(x)=r_2x^2+r_1x+r_0$，有 r_0，r_1，$r_2\in\{0,1\}$，因此全体剩余类为 $\{0,1,x,x+1,x^2,x^2+1,x^2+x,x^2+x+1\}$。

有 $((x+1)+(x^2+x))[\bmod\ p(x)]=(x+1+x^2+x)[\bmod\ p(x)]=x^2+1$

$$((x+1)\cdot(x^2+x))[\bmod\ p(x)]=(x^3+x)[\bmod\ p(x)]=1$$

定理 7.5 任何首一多项式可分解为首一即约多项式之积：

$$f(x)=p_1(x)^{r_1}p_2(x)^{r_2}\cdots p_n(x)^{r_n}$$

式中，r_i 为正整数，$p_1(x)$，$p_2(x)$，…，$p_n(x)$为 $f(x)$的因式，且是 n 个互不相同的即约多项式，当不考虑因式的顺序时，这种分解是唯一的。

定理 7.6 一定存在多项式 $m(x)$，$n(x)$，使 $m(x)\cdot a(x)+n(x)\cdot b(x)=(a(x),b(x))$，$(a(x),b(x))$为多项式 $a(x)$，$b(x)$的最大公因式。

7.1.3 线性空间

（1）线性空间

如果定义在 F 上的 n 重集合 $\boldsymbol{V}$ 对任何 u，$v\in\boldsymbol{V}$，c，$d\in F$，满足下列条件：

① 存在单位元 $e\in\boldsymbol{V}$，有 $e+v=v+e=v$；

② 有逆元 $v^{-1}\in\boldsymbol{V}$，使 $v+v^{-1}=v^{-1}+v=e$；

③ 满足封闭性、交换律，有 $cu+dv=dv+cu\in\boldsymbol{V}$；

④ 结合律、分配律成立，有 $c(u+v)=cu+cv$，$(c+d)v=cv+dv$。

则称 $\boldsymbol{V}$ 是 F 上的线性空间或矢量空间。

【例 7.4】 全体 n 重矢量的集合$\{a_1a_2\cdots a_n|a_i\in\{0,1\}\}$构成一个线性空间。

系数为实数 R 且次数少于 n 的全体多项式$\{f_{n-1}x^{n-1}+f_{n-2}x^{n-2}+\cdots+f_1x+f_0\mid f_i\in R\}$组成一个线性空间。

元素是实数的 $m\times n$ 矩阵也是一个线性空间（矩阵的每一行或每一列可看成一个矢量）。

（2）线性组合

F 上的矢量 $\boldsymbol{v}_1$，$\boldsymbol{v}_2$，…，$\boldsymbol{v}_k$ 的线性组合为

$$\boldsymbol{u}=b_1\boldsymbol{v}_1+b_2\boldsymbol{v}_2+\cdots+b_k\boldsymbol{v}_k\qquad b_i\in F$$

例如，线性空间 $\boldsymbol{V}=\{a_1a_2a_3|a_i\in\{0,1,2\}\}$，设 $\boldsymbol{v}_1=012$，$\boldsymbol{v}_2=101$，$\boldsymbol{v}_3=221$，$b_1=2$，$b_2=0$，$b_3=1$，则 $\boldsymbol{u}=2(012)+0(101)+1(221)=(021)+(000)+(221)=212$。

（3）线性相关

$\boldsymbol{v}_1$，$\boldsymbol{v}_2$，…，$\boldsymbol{v}_k$ 是线性空间 $\boldsymbol{V}$ 中一组非零矢量，若存在一组不全为 0 的标量 b_1，b_2，…，b_k，$b_i\in F$，使 $b_1\boldsymbol{v}_1+b_2\boldsymbol{v}_2+\cdots+b_k\boldsymbol{v}_k=0$，则称这组矢量线性相关。

（4）线性无关

$\boldsymbol{v_1}$，$\boldsymbol{v_2}$，…，$\boldsymbol{v_k}$为 k 个非零的矢量，只有当 k 个标量全为 0，即 $b_1 = b_2 = \cdots = b_k = 0$ 时，等式 $b_1\boldsymbol{v_1} + b_2\boldsymbol{v_2} + \cdots + b_k\boldsymbol{v_k} = 0$ 才成立，则称这 k 个矢量线性无关。

例如，{0，1}上的 4 重矢量（0001），（0010），（0100），（1000）线性无关，而（1001），（0010），（0100），（1111）线性相关，因为（1001）+（0010）+（0100）+（1111）= 0000。

（5）张成

线性空间 $\boldsymbol{V}$ 中的每一个非零矢量，如果可以由其中的一组矢量集合 $\boldsymbol{S}$ 中的矢量线性组合而成，则称 $\boldsymbol{S}$ 张成 $\boldsymbol{V}$。

比如{0，1}上的矢量（001），（010），（011），（100），（101），（110），（111）可由（001），（010），（100）张成，也可由（011），（100），（101）张成。

- 基底：张成整个线性空间的一组线性无关矢量。
- 维数：基底中线性无关矢量的数目。

如果 $\boldsymbol{V}$ 是 k 维线性空间，则 $\boldsymbol{V}$ 中任何 k 个线性无关矢量都是 $\boldsymbol{V}$ 的基底。

例如，（001），（010），（100）是{0，1}上 3 重矢量空间的基底，同样（001），（010），（101）也可作为其基底，另外（011），（101），（100）也是基底。该线性空间的维数是 3。

n 维线性空间的自然基底是（00…01），（00…10），（01…00），（10…00）。

7.2 群 与 环

7.2.1 群的定义

1. 群的定义

群 G 是一些元素构成的集合，该集合中定义一种运算“$*$”（加法或乘法），满足

（1）封闭性，对任何 $a, b \in G$，有 $a * b \in G$；

（2）结合律，对任何 $a, b, c \in G$，有$(a * b) * c = a * (b * c)$；

（3）存在单位元 $e \in G$，使对任何 $a \in G$ 有 $a * e = e * a = a$；

（4）对任何 $a \in G$ 有逆元 $a^{-1} \in G$，使 $a * a^{-1} = a^{-1} * a = e$。

习惯上，若群的运算是加法，则简称加群；若为乘法则简称为乘群。

【例 7.5】 整数集合对加法运算很明显满足封闭性和结合律，任何整数加 0 等于其自身，故加法单位元为 0，任意一个整数 x 的逆元是其相反数$-x$，因此可判断全体整数构成加群。

类似地，全体偶数、实数、复数也构成加群。

另外，n 阶方阵对加法运算也构成群，单位元是零矩阵。

但对乘法来说，整数集合虽然满足封闭性和结合律，而且乘法单位元为 1，但是由于除 1 和−1 外，其他元素均无逆元，所以整数集合不能构成乘群。同样，因为元素 0 无逆元，故全体实数或复数集合也不能构成乘群。但如果把 0 排除掉，非 0 实数和非 0 复数集合在乘法运算下都是群，乘法单位元是 1，元素 x 的逆元为 $\frac{1}{x}$。

- **交换群：**如果“$*$”运算还满足交换律，即对任何 $a, b \in G$，有 $a * b = b * a$，则 G 称为交换群。

加群是交换群，而乘群不一定是交换群，如矩阵乘法不满足交换律。

- **群的阶**：群的阶就是群中所含元素的个数。

整数加群和非 0 实数乘群的阶都是无穷值。

- **有限群**：阶为有限值的群称为有限群。

【例 7.6】 模 d 的全体剩余类{0, 1, … , d−1}对模 d 加法运算如表 7-1 所示。

表 7-1 模 d 全体剩余类对模 d 的加法运算表

+	0	1	…	d–2	d–1
0	**0**	1	…	d–2	d–1
1	1	2	…	d–1	**0**
…			…		
d–2	d–2	d–1	…	d–4	d–3
d–1	d–1	**0**	…	d–3	d–2

从表 7-1 看出，模 d 的全体剩余类对模 d 的加法运算满足封闭性、结合律和交换律，单位元为 0，0 的逆元为 0，元素 i 的逆元为 d–i，因此构成交换加群。该群的阶为 d，是有限群。

同样，以 $p(x)$为模的多项式的全体剩余类对模 $p\ (x)$的加法运算也构成交换加群。

【例 7.7】 集合{0，1}上的任意多项式以 $p\ (x) = x^3 + x + 1$ 为模，所得全体剩余类为{0, 1, x, $x + 1$，x^2，$x^2 + 1$，$x^2 + x$, $x^2 + x + 1$}，该剩余类对模 $p\ (x)$的加法运算如表 7-2 所示。

表 7-2 模 $p\ (x) = x^3 + x + 1$ 的加法表

+	0	1	x	$x + 1$	x^2	$x^2 + 1$	$x^2 + x$	$x^2 + x + 1$
0	**0**	1	x	$x + 1$	x^2	$x^2 + 1$	$x^2 + x$	$x^2 + x + 1$
1	1	**0**	$x + 1$	x	$x^2 + 1$	x^2	$x^2 + x + 1$	$x^2 + x$
x	x	$x + 1$	**0**	1	$x^2 + x$	$x^2 + x + 1$	x^2	$x^2 + 1$
$x + 1$	$x + 1$	x	1	**0**	$x^2 + x + 1$	$x^2 + x$	$x^2 + 1$	x^2
x^2	x^2	$x^2 + 1$	$x^2 + x$	$x^2 + x + 1$	**0**	1	x	$x + 1$
$x^2 + 1$	$x^2 + 1$	x^2	$x^2 + x + 1$	$x^2 + x$	1	**0**	$x + 1$	x
$x^2 + x$	$x^2 + x$	$x^2 + x + 1$	x^2	$x^2 + 1$	x	$x + 1$	**0**	1
$x^2 + x + 1$	$x^2 + x + 1$	$x^2 + x$	$x^2 + 1$	x^2	$x + 1$	x	1	**0**

表中加法单位元为 0，每个元素的逆元为其自身。

【例 7.8】 模 6 的非 0 剩余类{1, 2, 3, 4, 5} 对模 6 乘法运算如表 7-3 所示。

表 7-3 模 6 的非 0 余数全体对模 6 的乘法运算表

×	1	2	3	4	5
1	**1**	2	3	4	5
2	2	4	0	2	4
3	3	0	3	0	3
4	4	2	0	4	2
5	5	4	3	2	**1**

从表 7-3 看出，模 6 的非 0 剩余类对模 6 乘法运算中，单位元为 1，元素 2, 3, 4 无逆元，因此不构成群。

2. 循环群

由元素α 的一切幂次所构成的群 $\{\alpha^0 = e, \alpha, \alpha^2, \cdots, \alpha^{n-1} \mid \alpha^n = e\}$ 称为循环群。元素α 称为循环群的生成元。使$\alpha^n = e$ 的最小正整数 n 称为元素α 的阶。

定义中幂次是相对于乘群而言的，同样，可构成循环加群，循环加群$\{e, \alpha, 2\alpha, \cdots, (n-1)\alpha \mid n\alpha = e\}$由生成元$\alpha$ 的一切倍次构成。循环加群中元素α 的阶是使 $n\alpha = e$ 的最小正整数 n。

由于$\alpha^m \cdot \alpha^n = \alpha^{m+n} = \alpha^n \cdot \alpha^m$，所以循环群都是交换群。

定理 7.7　交换群中的每一个元素α 都能生成一个循环群，元素α 的阶就是循环群的阶。

【例 7.9】　模 9 的全体剩余类{0, 1, 2, 3, 4, 5, 6, 7, 8}在模 9 加法运算“*”下构成群，取生成元为 2，则

$\alpha = 2$　　$2\alpha = 2 * 2 = 4$　　$3\alpha = 2\,\alpha * \alpha = 4 * 2 = 6$

$4\alpha = 3\,\alpha * \alpha = 6 * 2 = 8$　　$5\alpha = 4\,\alpha * \alpha = 8 * 2 = 1$

$6\alpha = 5\,\alpha * \alpha = 1 * 2 = 3$　　$7\alpha = 6\,\alpha * \alpha = 3 * 2 = 5$

$8\alpha = 7\,\alpha * \alpha = 5 * 2 = 7$　　$9\alpha = 8\,\alpha * \alpha = 7 * 2 = 0 = e$

因而由 2 生成的循环加群为{0，2，4，6，8，1，3，5，7}，该循环群和生成元 2 的阶都为 9。

再取生成元为 6，则　　$6 * 6 = 3$

$6 * 6 * 6 = 3 * 6 = 0$

$6 * 6 * 6 * 6 = 0 * 6 = 6$

生成的循环加群为{3, 0, 6}，该循环群和生成元 6 的阶都为 3。

元素阶的性质：

（1）若 a 是 n 阶元素，则 $a^m = e$（对于加法为 $ma = e$）的充要条件是 n 整除 m。

（2）若某一群中，a 为 n 阶元素，b 为 m 阶元素，且$(n, m) = 1$，则元素 $a \cdot b$（或 $a + b$）的阶为 $n \cdot m$。

（3）若 a 为 n 阶元素，则元素 a^k（或 ka）的阶为 $\dfrac{n}{(n,k)}$。

7.2.2　子群

1. 子群的定义

若群 G 的非空子集 G' 对于 G 中所定义的代数运算也构成群，则称 G' 为 G 的子群。

例如，偶数加群是整数加群的子群。一般来说，某一整数 m 的所有倍数所构成的集合是整数加群的子群。

定理 7.8　有限群的子群的阶一定整除群的阶。

例 7.9 中生成元 6 在模 9 加法运算下构成的群 $G' = \{0，3，6\}$是全体剩余类群 $G = \{0，1，2，3，4，5，6，7，8\}$的一个子群。$G'$ 的阶 3 整除 G 的阶 9。

2. 群的陪集分解

设 G' 为群 G 的子群，取 $h \in G$，则称 $h * G'$ 为 G' 的左陪集，称 $G' * h$ 为 G' 的右陪集。当

G 是交换群时，子群 G' 的左、右陪集是相等的，元素 h 称为陪集首。

设 $G'=\{g_1, g_2, \cdots, g_n\}$，$G'$ 的阶为 n ，又设 G' 为群 G 的子群，由定理 7.8，设 G 的阶为 $n\cdot m$，那么可将 G 完备地分成 m 个陪集（子群本身也是一个陪集），如表 7-4 所示。

表 7-4　陪集分解表

陪　　集					说　　明
$h_1*g_1=g_1=e$	g_2	$\cdots$	g_{n-1}	g_n	陪集首 $h_1=e$，子群 G'
$h_2*g_1=h_2$	h_2*g_2	$\cdots$	h_2*g_{n-1}	h_2*g_n	陪集首 h_2，陪集 h_2*G'
		⋮			⋮
$h_{m-1}*g_1=h_{m-1}$	$h_{m-1}*g_2$	$\cdots$	$h_{m-1}*g_{n-1}$	$h_{m-1}*g_n$	陪集首 h_{m-1}，陪集 $h_{m-1}*G'$
$h_m*g_1=h_m$	h_m*g_2	$\cdots$	h_m*g_{n-1}	h_m*g_n	陪集首 h_m，陪集 h_m*G'

陪集首的选择应注意：

（1）若陪集首 h 是子群 G' 中的元素，则陪集 $h*G'$ 与子群 G' 相同。

（2）若陪集首 h 不是子群 G'中的元素，则陪集 $h*G'$ 与子群 G' 相交为空集。

（3）若陪集首 h_j 不是陪集 h_i*G'中的元素，则两陪集 h_i*G' 与 h_j*G' 相交为空集。

（4）陪集 $h*G'$ 中的每一个元素都可作为其陪集首 h，陪集元素不变，仅排列顺序改变。

由以上性质可知，两个陪集要么相等要么不相交。为使群的分解完备，应选择前面未出现过的元素作为当前陪集的陪集首，这样，整个群将分解成若干个互不相交的陪集，无一遗漏，无一重叠。

【例 7.10】　模 9 加法运算下 $G'=\{0, 3, 6\}$是 $G=\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$的子群，取 $h=0$，则 $0+G'=\{0, 3, 6\}=G'$。

若取 $h=1$，则 $1+G'=\{1, 4, 7\}$，$(1+G')\cap G'=\varnothing$；

若取 $h=2$，则 $2+G'=\{2, 5, 8\}$，$(2+G')\cap G'=\varnothing$，$(1+G')\cap(2+G')=\varnothing$；

若取 $h=3$，则 $3+G'=\{3, 6, 0\}=G'$。

因此可将 G 分解为 3 个陪集，即

$$\begin{aligned} G'&: \quad 0 \quad 3 \quad 6 \\ 1+G'&: \quad 1 \quad 4 \quad 7 \\ 2+G'&: \quad 2 \quad 5 \quad 8 \end{aligned}$$

7.2.3　环

1. 环的定义

环 F 是一些元素构成的集合，该集合中定义加法和乘法两种运算，满足

（1）对加法是一个交换群，即满足封闭性、结合律、交换律、存在加法单位元和逆元。

（2）对乘法具有封闭性和结合律。

（3）分配律成立，即对任何 $a, b, c\in F$，有

$$a(b+c)=ab+ac \qquad (a+b)c=ac+bc$$

可以看出，环对乘法运算不要求单位元和逆元。

如果集合 F 对乘法运算还满足交换律，则称为交换环。

【例 7.11】　全体整数构成交换环。

n 阶非奇异方阵构成环。

系数取自实数集合的全体多项式构成交换环。

2. 剩余类环

剩余类环是一类重要的环，它是构成有限域的基础。

以整数 d 为模进行除法运算所得的全体剩余类可构成环，称为整数剩余类环。

如表 7-1 所示，模 d 的全体剩余类 $F = \{0, 1, \cdots, d-1\}$对模 d 加法运算构成交换加群；显然集合 F 对模 d 乘法运算满足封闭性、结合律和交换律；对于集合 F 来说，有

$$(a + b) \text{ c } [\text{mod } \ d] = (ac + bc) \ \ [\text{mod } \ d]$$

$$a\,(b + c)\,[\text{mod } \ d] = (ab + ac) \ \ [\text{mod } \ d]$$

说明集合 F 还满足分配律，因此模 d 的全体剩余类构成交换环。

如模 4 的剩余类环{0 , 1 , 2 , 3}、模 5 的剩余类环{0 , 1 , 2 , 3 , 4}等。

同理，模 $p\,(x)$的全体剩余类对模 $p\,(x)$的运算构成交换环，称为多项式剩余类环。

3. 子环

设 S 是环 F 的一个非空子集，若 S 关于 F 的代数运算也构成一个环，则称 S 是 F 的子环，F 是 S 的扩环。

【例 7.12】　某一整数 m 的全体倍数构成交换环{0 , ±m , ±2m , ±3m , … }，且是整数环的一个子环。

$F_2[\,x\,]$是集合{0，1}上的多项式环，则多项式 $f\,(x) = x^2 + 1$ 的倍式可构成 $F_2[\,x\,]$的子环，为

$$S_f[\,x\,] = \{0\,,\, x^2 + 1\,,\, x\,(x^2 + 1)\,,\, (x + 1)\,(x^2 + 1)\,,\, \cdots\}$$

在交换环 F 中，由一个元素 $a \in F$ 的所有倍数及其线性组合而生成的子环 $S_a = \{xa + na \mid x \in F\,,\, n \in Z\}$中，元素 a 为该子环的生成元。当 a 为多项式时，也称为生成多项式。

在例 7.12 中，m 和 $f\,(x)$ 分别是子环的生成元和生成多项式。

7.3　域

7.3.1　域的定义

域是一些元素构成的集合，该集合中定义加法和乘法两种运算，满足

（1）对加法构成交换加群；

（2）全体非零元素对乘法构成交换乘群；

（3）加法和乘法间具有分配律。

例如，全体复数和全体实数都构成域，全体整数则不能构成域。

- **域的阶**：域中元素的个数。
- **有限域**：元素个数有限的域，用 GF(q)表示 q 阶有限域。

纠错码理论中经常需要将多项式 x^n-1（特别是 $n = 2^m-1$）因式分解，而由剩余类构成的有限域是多项式 x^n-1 因式分解的理论基础，下面将逐一介绍由整数剩余类环和多项式剩余类环引出的两种重要的有限域。

定理 7.9　设 d 为素数，则以 d 为模的整数剩余类环构成 d 阶有限域 GF(d)。

证明：域比环多三个条件，即乘法满足交换律、存在乘法单位元、非零元素有乘法逆元。

显然，模 d 剩余类环对模 d 乘法满足交换律，而且乘法单位元为 1，所以只要证明非零元素有乘法逆元即可。

设 r 为模 d 的余数，则 $0 \leqslant r < d$。

由于 d 为素数，又设 r' 为模 d 的非 0 余数，则 r' 与 d 一定互素，即（r' , d）= 1。

根据定理 7.3，存在整数 m，n，使 $mr' + nd =$（r' , d）= 1。

等式两边同时对 d 取模，则 $m \cdot r' \ [\mathrm{mod}\ \ d\ ] = 1$。即 m 为 r' 的模 d 乘法逆元，说明以素数 d 为模的任一非 0 余数的逆元存在。又由于满足 $0 \leqslant r < d$ 的整数 r 共有 d 项，所以以素数 d 为模的整数剩余类环构成 d 阶有限域。

【例 7.13】　$d = 2$ 构成域 GF (2) = {0 , 1}，$d = 5$ 构成域 GF (5) = {0 , 1 , 2 , 3 , 4}。

由表 7-1 已知以任意整数 d 为模的全体剩余类对模 d 加法构成交换加群，因此这里只列出乘法运算。表 7-5（a）和表 7-5（b）分别为模 2 乘法表和模 5 乘法表。

表 7-5　乘法表

（a）模 2 乘法表

×	1
1	

注：乘法不考虑元素 0，
乘法单位元为 1，1 的逆元为 1。

（b）模 5 乘法表

×	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

注：1 为乘法单位元，1 的逆元为 1，2 与 3 互逆，
4 的逆元为 4。

而 $d = 6$ 时，由于 6 不是素数，由例 7.8 已知，模 6 的非 0 剩余类集合{ 1 , 2 , 3 , 4 , 5}对模 6 乘法运算不能构成交换群，因而模 6 的全体剩余类{ 0 , 1 , 2 , 3 , 4 , 5}也就不能构成域。

定理 7.10　设 $p(x)$为系数取自 GF (q)上的 n 次即约多项式，则以 $p(x)$为模的多项式剩余类环构成 q^n 阶有限域 GF (q^n)。

【例 7.14】　系数取自 GF (2) = {0 , 1}的全体多项式集合用模 $p(x) = x^3 + x + 1$ 除，所得余式构成有限域 GF (2^3) = {0 , 1 , x , $x + 1$, x^2 , $x^2 + 1$, $x^2 + x$, $x^2 + x + 1$}，其加法运算见表 7-2，乘法表见表 7-6。

表 7-6　模 $p\ (x) = x^3 + x + 1$ 乘法表

×	1	x	$x + 1$	x^2	$x^2 + 1$	$x^2 + x$	$x^2 + x + 1$
1	**1**	x	$x + 1$	x^2	$x^2 + 1$	$x^2 + x$	$x^2 + x + 1$
x	x	x^2	$x^2 + x$	$x + 1$	**1**	$x^2 + x + 1$	$x^2 + 1$
$x + 1$	$x + 1$	$x^2 + x$	$x^2 + 1$	$x^2 + x + 1$	x^2	**1**	x
x^2	x^2	$x + 1$	$x^2 + x + 1$	$x^2 + x$	x	$x^2 + 1$	**1**
$x^2 + 1$	$x^2 + 1$	**1**	x^2	x	$x^2 + x + 1$	$x + 1$	$x^2 + x$
$x^2 + x$	$x^2 + x$	$x^2 + x + 1$	**1**	$x^2 + 1$	$x + 1$	x	x^2
$x^2 + x + 1$	$x^2 + x + 1$	$x^2 + 1$	x	**1**	$x^2 + x$	x^2	$x + 1$

注：乘法单位元为 1，每个元素都有逆元；1 的逆元为其自身 1；x 与 x^2+1 互为逆元；$x + 1$ 与 $x^2 + x$ 互为逆元；
x^2 与 $x^2 + x + 1$ 互为逆元。

令 GF (Q) =GF (q^n)，称 GF (Q)为 GF (q)的扩展域，GF (q)为 GF (Q)的子域。对于编码来说，扩展域更有用，用扩展域构成的码有更长的码长，有更多的码矢用来代表更多的消息。

定理 7.11 有限域的阶必为其子域阶之幂，即 $Q = q^n$。

7.3.2 有限域的本原元

域中全体非零元素构成交换乘群，由定理 7.7 可知，该乘群中每一个元素都能生成循环群，但各元素阶不一定相等。

- **本原元**：在 GF (q)中，某一元素α 的阶为 $q-1$，即$\alpha^{q-1} = e$（$q-1$ 是使等式成立的最小正整数），则称α 为本原元。
- **本原元多项式**：是以本原元为根的即约多项式。

【例 7.15】 集合{0 , 1}上模多项式 $p(x) = x^3 + x + 1$ 的全体剩余类在模 $p(x)$的运算下构成域 GF (8) = {0 , 1 , x , $x+1$, x^2 , x^2+1 , x^2+x , x^2+x+1}。

若元素α 有$\alpha^3 = \alpha + 1$，则$\alpha^7 = \alpha^3 \cdot \alpha^3 \cdot \alpha = (\alpha + 1)(\alpha + 1)\ \alpha = \alpha^3 + \alpha = 1$，集合{1 , α, α^2, α^3, α^4, α^5, α^6 }是一个循环乘群，循环群一定是交换群，而集合{0 , 1 , α, α^2, α^3, α^4, α^5, α^6 }根据表达式$\alpha^3 = \alpha + 1$，如表 7-7 所示，构成交换加群。（单位元为 0；各个元素的逆元为其自身；分配律显然是成立的。）因此集合{0 , 1 , α, α^2, α^3, α^4, α^5, α^6 }也构成域，记为 GF′ (8)。

表 7-7 {0 , 1 , α, α^2, α^3, α^4, α^5, α^6 }的加法表

+	0	1	α	α^2	α^3	α^4	α^5	α^6
0	**0**	1	α	α^2	α^3	α^4	α^5	α^6
1	1	**0**	α^3	α^6	α	α^5	α^4	α^2
α	α	α^3	**0**	α^4	1	α^2	α^6	α^5
α^2	α^2	α^6	α^4	**0**	α^5	α	α^3	1
α^3	α^3	α	1	α^5	**0**	α^6	α^2	α^4
α^4	α^4	α^5	α^2	α	α^6	**0**	1	α^3
α^5	α^5	α^4	α^6	α^3	α^2	1	**0**	α
α^6	α^6	α^2	α^5	1	α^4	α^3	α	**0**

令 $x = x \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha$，则

$x^2 = x^2 \ [\mathrm{mod}\ \ p(x)] = x \ [\mathrm{mod}\ \ p(x)] \cdot x \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha \cdot \alpha = \alpha^2$；

$x + 1 = x^3 \ [\mathrm{mod}\ \ p(x)] = x \ [\mathrm{mod}\ \ p(x)] \cdot x^2 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha \cdot \alpha^2 = \alpha^3$；

$x^2 + x = x^4 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha^4$；

$x^2 + x + 1 = x^5 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha^5$；

$x^2 + 1 = x^6 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha^6$；

$0 \rightarrow 0$；

$1 \rightarrow 1$；

$x^2 \cdot (x^2 + x) = x^2 + 1 = x^2 \cdot x^4 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha^2 \cdot \alpha^4 = \alpha^6$；

$x^2 \cdot (x^2 + x + 1) = 1 = x^2 \cdot x^5 \ [\mathrm{mod}\ \ p(x)] \rightarrow \alpha^2 \cdot \alpha^5 = 1$；…

GF (8)与 GF′(8) 的元素之间存在一一对应关系，并且这两个集合的所有代数性质也都一一对应。数学上把这样的系统看成本质上完全相同，研究其中一个也就代替了对另一个的研究，因此可根据各自不同的特点，将各个系统应用在不同的情况下。

在例 7.15 中，由于$\alpha^{q-1}=1$，所以α 为本原元。由于有对应关系，故 GF (8)中的所有非零元素都可表示成本原元α 的方幂：$\alpha, \alpha^2, \cdots, \alpha^{q-2}, \alpha^{q-1}=e$。利用表达式$\alpha^3=\alpha+1$ 还可将幂次形式变换为α 的 2 次多项式，并且由于 n–1 次多项式与 n 维矢量一一对应，故也可表示为 3 维矢量，因此元素的表示形式有剩余类、多项式（线性组合）、幂级数及矢量，如表 7-8 所列。

表 7-8　GF (2^3)中元素的四种表示

剩　余　类	线 性 组 合	幂　级　数	矢　　量
0	0	0	000
1	1	1	001
x	α	α	010
x^2	α^2	α^2	100
$x+1$	α + 1	α^3	011
x^2+x	$\alpha^2+\alpha$	α^4	110
x^2+x+1	$\alpha^2+\alpha+1$	α^5	111
x^2+1	α^2+1	α^6	101

定理 7.12　GF (q)的所有元素是方程 $x^q-x=0$ 的根，反之，方程 $x^q-x=0$ 的根必在 GF(q)中。

证明：方程 $x^q-x=0$ 共有 q 个根，而 $x^q-x=x(x^{q-1}-1)$，显然 0 元素是方程 $x^q-x=0$ 的一个根，下面将寻找该方程的其他 $q-1$ 个根，也就是寻找方程 $x^{q-1}-1=0$ 的 $q-1$ 个根。

GF(q)中的 $q-1$ 个非零元素都可以表示成本原元α 的方幂α^i 的形式：$\{\alpha^0, \alpha^1, \alpha^2, \cdots, \alpha^{q-2}\}$，由于$\alpha$为本原元，有$\alpha^{q-1}=1$，则每一个非零元素均满足$(\alpha^j)^{q-1}=(\alpha^{q-1})^i=1^i=1$。

即
$$(\alpha^j)^{q-1}-1=0$$

说明 GF (q)中的 $q-1$ 个非零元素α^i（$0\leqslant i<q-1$）都是 $x^{q-1}-1=0$ 的根，也就是说，GF (q)中的全部元素都是 $x^q-x=0$ 的根，换句话说，$x^q-x=0$ 的全部根都在 GF (q)中。

根据这一定理，可将 x^q-x 在 GF (q)中完全分解成一次因式：$x^q-x=x\prod_{i=0}^{q-2}(x-\alpha^i)$。

7.3.3　有限域的结构

1. 有限域的特征

- **有限域的特征**：是有限域中乘法单位元 e 关于加法的级，也就是使 $p\cdot e=0$ 的最小正整数 p。

  GF (2) = {0 , 1}中乘法单位元 $e=1$，有 $1+1=0$，所以 $p=2$。

定理 7.13　有限域的特征必为素数。

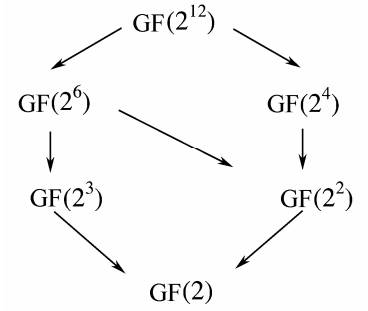
- **素域**：是 GF (q)的最小子域，表示为 GF (p) = {0 , e , 2e , ⋯ , (p−1)e}。

  整数剩余类环构成的域一定是素域。

定理 7.14　有限域的阶必为其特征之幂，即 $q=p^m$。

结合定理 7.11，可将有限域逐步分解成各个子域，直至素域。

【例 7.16】　将 GF (2^{12})分解成子域。



定理 7.15　在以 p 为特征的域 GF (q)中，对于任意α，$\beta \in$ GF (q)，恒有

$$(\alpha+\beta)^p=\alpha^p+\beta^p$$

推论 1　若β_1 , β_2 ,…, β_k 是以 p 为特征的域中的元素，则对任意正整数 n 恒有

$$(\sum_{i=1}^{k}\beta_i)^{p^n}=\sum_{i=1}^{k}\beta_i^{p^n}$$

2. 有限域的共轭根组

前面介绍了多项式 $x^{p^m}-x$ 在 GF (p^m)域上的因式分解。由于多项式是否即约与它所在的域密切相关，下面将通过共轭根组的介绍引出多项式 $x^{p^m}-x$ 在 GF (p)中的因式分解。

定理 7.16　对 GF (p^m)中的任意元素β，恒有 $\beta^{p^m}=\beta$ 。

定理 7.17　设 f (x)是系数取自 GF (p)的 k 次即约多项式，$\beta \in$ GF (p^m)，若β 是 f (x)的根，则 β^{p^r} （0≤r＜k）也是 f (x)的根。

证明：设 $f(x)=\sum_{i=0}^{k} f_i x^i$ ， $f_i \in$ GF (p)。

β 是 f (x)的根，结合推论 1 得

$$0=\sum_{i=0}^{k} f_i\beta^i=(\sum_{i=0}^{k} f_i\beta^i)^{p^r}=\sum_{i=0}^{k} f_i^{p^r}\beta^{ip^r}=\sum_{i=0}^{k} f_i^{p^r}(\beta^{p^r})^i$$

由于 $f_i \in$ GF (p)，又由推论 1 得

$$f_i^{p^r}=(f_i e)^{p^r}=(e+e+\cdots+e)^{p^r}=\sum_{j=1}^{f_i} e^{p^r}=\sum_{j=1}^{f_i} e=f_i$$

所以

$$\sum_{i=0}^{k} f_i(\beta^{p^r})^i=0$$

即 β^{p^r} 是 $\sum_{i=0}^{k} f_i x^i$ 的根，也就是 f (x)的根。

由于多项式 f(x)只有 k 个互不相同的根，这 k 个根就是 $\beta,\beta^p,\cdots,\beta^{p^{k-1}}$ （ $\beta^{p^k}=\beta$ ）。我们将这 k 个根称为 f (x)的共轭根组。

【例 7.17】　若元素$\beta \in$ GF (2^4)是 GF (2) = {0 , 1}上多项式 x^4+x+1 的根，寻找β 的所有共轭根。

解：因为β 是 x^4+x+1 的根，有$\beta^4+\beta+1=0$，得$\beta^4=\beta+1$。

将β^2代入多项式中，应用定理 7.15，有

$$(\beta^2)^4+\beta^2+1=(\beta^4)^2+\beta^2+1=(\beta+1)^2+\beta^2+1=\beta^2+1+\beta^2+1=0$$

同样将β^4，β^8分别代入多项式中有

$$(\beta^4)^4+\beta^4+1=(\beta+1)^4+\beta^4+1=\beta^4+1+\beta^4+1=0$$

$$(\beta^8)^4+\beta^8+1=(\beta^4)^8+\beta^8+1=(\beta+1)^8+\beta^8+1=\beta^8+1+\beta^8+1=0$$

说明β^2,β^4,β^8也都是多项式x^4+x+1的根，而由定理 7.16 可知$\beta^{16}=\beta$，因此共轭根组为 $\{\beta,\beta^2,\beta^4,\beta^8\}$。

再如，若γ是x^2+x+1的根，则$\gamma^2=\gamma+1$。

$$(\gamma^2)^2+\gamma^2+1=(\gamma+1)^2+\gamma^2+1=\gamma^2+1+\gamma^2+1=0$$

而$\gamma^4=\gamma$，所以共轭根组为 $\{\gamma,\gamma^2\}$。

- **最小多项式**：系数取自 GF (p)，且以$\beta\in$ GF (p^m)为根的所有首一多项式中，必有一个次数最低的多项式，称为β 的最小多项式。

最小多项式的性质：

① 最小多项式在 GF (p)上是即约的；

② 每一$\beta\in$ GF (p^m)，必有唯一的最小多项式；

③ β 的最小多项式能整除任何以β 为根的多项式，例如能整除多项式$x^{p^m}-x$。

根据定理 7.17，可在域 GF (p^m)上将β 的最小多项式表示为 $m(x)=\prod_{i=0}^{k-1}(x-\beta^{p^i})$。

上式表明：共轭根组内的所有元素共享同一最小多项式。

由于 m (x)在 GF (p)上即约，故结合定理 7.12，得出以下推论。

推论 2　设 $m_1(x)$, $m_2(x)$, … , $m_t(x)$为 GF (p^m)中各元素的最小多项式，那么可将多项式$x^{p^m}-x$在 GF (p)上分解为 $x^{p^m}-x=m_1(x)m_2(x)\cdots m_t(x)$。

至此，我们已经可以求解多项式$x^{p^m}-x$在 GF (p)上的即约因式了。

【例 7.18】　在 GF (2) = {0 , 1}系数域上，以 $p(x)=x^4+x+1$ 为模构成有限域 GF (2^4)，在 GF (2)上分解多项式$x^{16}-x$。

解：（1）由于 GF (2) = {0 , 1}，$e=1$，$1+1=0$，所以特征 $p=2$。

（2）寻找本原元。

设α为 $p(x)$的根，则 $\alpha^4=\alpha+1$。

$$\alpha^{15}=\alpha^4\alpha^4\alpha^4\alpha^3=(\alpha+1)(\alpha+1)(\alpha+1)\alpha^3=(\alpha^2+1)(\alpha+1+\alpha^3)$$

$$=\alpha^2+\alpha^5+\alpha+1=\alpha^2+(\alpha^2+\alpha)+\alpha+1=1$$

$\alpha^{15}=1$，因此α 为本原元，$p(x)$为本原多项式，GF (2^4)的 15 个非 0 元素都可以表示成α 的方幂α^0, α^1, … , α^{14}，如表 7-9 所示。

表 7-9　GF (2^4)中元素的四种表示

剩　余　类	线 性 组 合	幂　级　数	矢　　量
0	0	0	0000
1	1	1	0001
x	α	α	0010
x^2	α^2	α^2	0100
x^3	α^3	α^3	1000

续表

剩 余 类	线 性 组 合	幂 级 数	矢 量
$x+1$	$\alpha+1$	α^4	0011
x^2+x	$\alpha^2+\alpha$	α^5	0110
x^3+x^2	$\alpha^3+\alpha^2$	α^6	1100
x^3+x+1	$\alpha^3+\alpha+1$	α^7	1011
x^2+1	α^2+1	α^8	0101
x^3+x	$\alpha^3+\alpha$	α^9	1010
x^2+x+1	$\alpha^2+\alpha+1$	α^{10}	0111
x^3+x^2+x	$\alpha^3+\alpha^2+\alpha$	α^{11}	1110
x^3+x^2+x+1	$\alpha^3+\alpha^2+\alpha+1$	α^{12}	1111
x^3+x^2+1	$\alpha^3+\alpha^2+1$	α^{13}	1101
x^3+1	α^3+1	α^{14}	1001

而 0 和这 15 个非 0 元素正好是方程 $x^{16}-x=0$ 的 16 个根。

所以 $$x^{16}-x=x(x-\alpha^0)(x-\alpha^1)\cdots(x-\alpha^{14})$$

（3）按照定理 7.17，找出各个共轭根组，并构成相应的最小多项式，最小多项式的下标以共轭根组中元素的最低幂次表示。

$\{0\}$ $m(x)=x-0=x$

$\{\alpha^0\}$ $m_0(x)=x-\alpha^0=x+1$

$\{\alpha,\alpha^2,\alpha^4,\alpha^8\}$ $m_1(x)=(x-\alpha)(x-\alpha^2)(x-\alpha^4)(x-\alpha^8)$

$\{\alpha^3,\alpha^6,\alpha^{12},\alpha^9\}$ $m_3(x)=(x-\alpha^3)(x-\alpha^6)(x-\alpha^{12})(x-\alpha^9)$

$\{\alpha^5,\alpha^{10}\}$ $m_5(x)=(x-\alpha^5)(x-\alpha^{10})$

$\{\alpha^7,\alpha^{14},\alpha^{13},\alpha^{11}\}$ $m_7(x)=(x-\alpha^7)(x-\alpha^{14})(x-\alpha^{13})(x-\alpha^{11})$

（4）利用本原多项式$\alpha^4=\alpha+1$，将最小多项式化简。

$$\begin{aligned}m_1(x)&=(x-\alpha)(x-\alpha^2)(x-\alpha^4)(x-\alpha^8)=[x^2-(\alpha+\alpha^2)x+\alpha^3][x^2-(\alpha^4+\alpha^8)x+\alpha^{12}]\\&=x^4-[\alpha^4+\alpha^8+\alpha+\alpha^2]x^3+[\alpha^{12}+\alpha^3+(\alpha+\alpha^2)(\alpha^4+\alpha^8)]x^2\\&\quad-[\alpha^{12}(\alpha+\alpha^2)+\alpha^3(\alpha^4+\alpha^8)]x+\alpha^{15}\\&=x^4+x+1\end{aligned}$$

同理得 $$m_3(x)=x^4+x^3+x^2+x+1$$

$$m_5(x)=x^2+x+1$$

$$m_7(x)=x^4+x^3+1$$

（5）将 $x^{16}-x$ 因式分解

$$\begin{aligned}x^{16}-x&=m(x)\,m_0(x)\,m_1(x)\,m_3(x)\,m_5(x)\,m_7(x)\\&=x(x+1)(x^4+x+1)(x^4+x^3+x^2+x+1)(x^2+x+1)(x^4+x^3+1)\end{aligned}$$

（6）根据$\alpha^{15}=1$ 以及元素阶的定义及性质，可得元素 1 的阶为 1；$\alpha,\alpha^2,\alpha^4,\alpha^8,\alpha^7,\alpha^{14},\alpha^{13},\alpha^{11}$ 的阶为 15；$\alpha^3,\alpha^6,\alpha^{12},\alpha^9$ 的阶为 5；α^5,α^{10} 的阶为 3。

因为阶为 $q-1$ 的元素为本原元，所以除α为本原元外，$\alpha^2,\alpha^4,\alpha^8,\alpha^7,\alpha^{14},\alpha^{13},\alpha^{11}$ 也都是本原元，其相应的两个最小多项式 x^4+x+1，x^4+x^3+1 即为本原多项式。

本 章 小 结

本章是后续纠错编码理论的代数基础，介绍的主要内容有：

（1）任何整数和多项式可表示为除法运算，余数相同的集合称为剩余类，每一个整数（多项式）可分解为素数（即约多项式）之积。

（2）群是一些元素在某种运算下构成的集合，满足封闭性、结合律、存在单位元和逆元。循环群由一个元素的所有幂次（倍次）构成。群可由其非空子群完备地分解为若干互不相交的陪集。

（3）环中元素对加法是交换群，对乘法满足封闭性、结合律和分配律。模 d 的全体剩余类对模 d 运算构成整数剩余类环，模 $p(x)$ 的全体剩余类对模 $p(x)$ 运算构成多项式剩余类环。

（4）域中元素对加法是交换群，非零元素对乘法是交换群，满足分配律。以素数（即约多项式）为模的剩余类环构成有限域。GF (q) 中的 q 个元素就是方程 $x^q - x = 0$ 的 q 个根。每一个共轭根组对应唯一的最小多项式。多项式 $x^{p^m} - x$ 可在 GF (p) 上因式分解为其所有最小多项式的乘积。

本章的内容是围绕多项式 $x^n - 1$ 的因式分解而逐步展开的，学习本章后应能熟练求解多项式 $x^{2^m - 1} - 1$ 的即约因式。

思考题与习题

7.1　全体非负整数集合能否构成加群和乘群？

7.2　全体二进制 4 重矢量的集合能否构成交换群？若能，找出其中一个 4 阶子群，并构造陪集表。

7.3　若群 G 的阶为素数，问有几个有限循环子群？

7.4　集合{1 , −1 , i , −i }对乘法和加法是否构成群？

7.5　集合{0 , 1 , 2 , 3 }在模 4 运算下能否构成乘群或加群？

7.6　基于 GF (2) 上的多项式 $p(x) = x^5 + x^2 + 1$，构造 GF (2^5) 的加法表和乘法表，找出本原多项式。

7.7　以 $p(x) = x^5 + x^2 + 1$ 为模将多项式 $x^{31} - 1$ 分解为 GF (2) 上的即约多项式之积。

7.8　根据本原多项式 $p(x) = x^3 + x + 1$，在 GF (2) 上对 $x^8 - x$ 作因式分解。

7.9　根据本原多项式 $p(x) = x^3 + x^2 + 1$，构成 GF (2^3) 的元素表，将每一个非零元素表示成幂级数、线性组合、模 $p(x)$ 的剩余类及二进制矢量四种形式，求出各元素的阶及本原元。

7.10　若本原多项式为 $p(x) = x^4 + x + 1$，本原元为 α ，求出 GF (2) 的 4 次扩展域中每一个非零元素的乘法逆元。

7.11　判断 GF (2) 的 6 次扩展域中共有多少个本原元，有多少个本原多项式？

7.12　模 8 的全体剩余类所构成的加群是否是循环群，若是，确定每个元素的级，找出所有生成元。

7.13　证明实数域上一切有逆的 n 阶方阵对矩阵乘法构成一个群。

7.14　模 $p(x) = x^3 + x^2 + x + 1$ 的全体剩余类能否构成域？为什么？

7.15　模 $p(x) = x^4 + x^3 + x^2 + 1$ 能否作为 GF (2)上 4 次扩展域的本原多项式？

7.16　在模 8 的剩余类环中找出所有子环。

7.17　证明有限域的特征必为素数。

第 8 章
线性分组码

内容提要

线性分组码同时具有信息位分组和校验位与信息位呈线性关系两种特性。目前，几乎所有得到实际应用的纠错码都是线性的。本章首先介绍有关纠错码的基本概念，然后重点论述线性分组码的定义及其编译码理论。并介绍线性分组码的纠检错能力。最后介绍一种典型的线性分组码——汉明码。

知识要点

线性分组码，生成矩阵，校验矩阵，最小距离，纠检错能力，伴随式，标准阵列。

教学建议

线性分组码是整个纠错码中很重要的一类码，也是讨论各类码的基础。它概念清楚，易于理解，而且能方便地引出各类码中广为采用的一些基本参数和名称的定义，因此本章的内容将涉及整个纠错码的基本知识。建议学时数为 8 学时，重点学习线性分组码的构成理论及其编译码方法。为了便于学生理解，讲授时应结合实例。

8.1 纠错码的基本概念

8.1.1 信道纠错编码

近年来，随着计算机、卫星通信及高速数据网的飞速发展，数据的交换、处理和存储技术得到了广泛的应用，人们对数据传输和存储系统的可靠性提出了越来越高的要求。因此，如何控制差错，提高数据传输和存储的可靠性，成为现代数字通信设计工作者所面临的重要课题。

1948 年，香农（C.E.Shannon）提出了关于在有扰信道中传输信号的重要理论——香农第二定理。该定理指出，在信息传输速率 R 小于信道容量 C 的条件下，当码长 $n\to\infty$时，总可以找到平均误码率 $\overline{p_e}\to 0$ 的码，定理的证明采用随即编码技术。香农第二定理虽然为提高数据传输的可靠性指出了一个方向，但并没有给出构造码的具体方法，因此这一定理只有理论指导意义。纠错编码就是后人沿着香农指明的可行方向为寻求有效而可靠的编码方法而发展起来的一门学科，经过半个多世纪的努力，目前已有了许多有效的编译码方法，并形成了一门新的技术——纠错编码技术。

这里所讲的纠错编码即信道编码，与本书前面讨论的信源编码一样，都是一种编码，但两者的侧重点是不同的。信源编码的侧重点是压缩冗余度或压缩熵率以得到信息的有效表示，提高信息的传输速率。信道编码的侧重点是提高信息传输时的抗干扰能力以增加信息传输的可靠性。

8.1.2 差错类型

下面讨论码字序列 $\boldsymbol{c}$ 通过离散信道时发生的情况。信道可分为无记忆信道和有记忆信道。

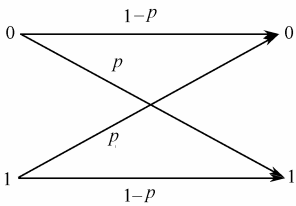


图 8-1　二进制对称信道

在无记忆信道中，噪声对传输码元的影响是相互独立的，即每一个差错的出现与其前后是否有错无关。图 8-1 所示的模型就是这种信道的一个例子。这里，1 错成 0 或 0 错成 1 的概率相等，均为 p。该模型描述的信道就是第 1 章介绍的二进制对称信道。在无记忆信道中，错误是随机产生的，因此被称为随机错误，无记忆信道也称为随机信道（random channel）；深空信道和卫星信道都属于随机信道的类型。

在有记忆信道中，各种干扰所造成的错误往往不是单个地，而是成群、成串地出现，也就是一个错误的出现，往往引起其前后码元的错误，表现出错误之间有相关性。如高频、有线信道及数据存储系统中磁带、磁盘、光盘或其他存储体的缺陷或读写头接触不良所引起的错误，都属于这种类型。图 8-2 就是这种信道的一个模型。图中，信道有两种状态，好状态 S_1 和坏状态 S_2，它们各自被描述为二进制对称信道，好状态工作时的误码率 p_1 远远小于坏状态工作时的误码率 p_2。由于 q_1□ q_2，信道经常工作在 S_1 状态，但偶而会工作在 S_2 状态，因此它产生突发错误。在一个突发错误持续期内，开头和末尾的码元总是错误的，中间的码元不一定都错，但错误的码元相对较多。

就实际信道而言，由于其干扰的复杂性，往往是两种错误并存。随机错误与突发错误并存的信道，称为组合信道或复合信道。

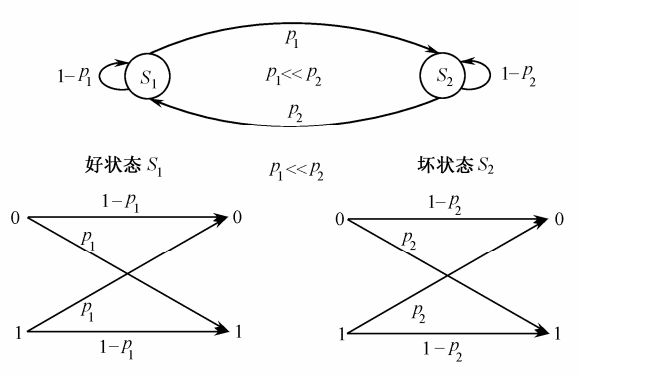


图 8-2　有记忆信道模型

针对上述不同类型的信道，必须设计出能纠正随机错误、突发错误，或既能纠正随机错误、又能纠正突发错误的码。

8.1.3　差错控制系统模型及分类

由于我们所关心的是信息传输的可靠性问题，为了方便研究，可将图 1.1 所示的信息传输系统模型简化成图 8-3 所示的简化模型。

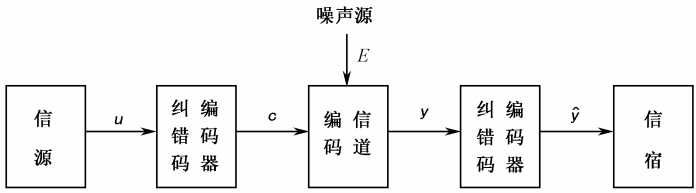


图 8-3　简化的信息传输系统模型

在图 8-3 中，等效信源是指原来的信源和信源编码器，它的输出是二进制信息序列 $\boldsymbol{u}$，纠错编码器按一定的编码规则将 $\boldsymbol{u}$ 编成码字 $\boldsymbol{c}$ 遣入信道，信道是包括调制器、传输媒质和解调器在内的数字信道（也称编码信道），它的输入是码字 $\boldsymbol{c}$，输出是接收码字 $\boldsymbol{y}$，纠错译码器根据编码规则对接收码字进行译码，输出估值 $\hat{\boldsymbol{y}}$，信宿包括信源译码器和用户，它的输入是经过纠错的估值序列 $\hat{\boldsymbol{y}}$。模型突出了以控制差错为目的的纠错编码器和纠错译码器，因此也称为差错控制系统。

目前在差错控制系统中使用的码按其纠错能力的不同可分为两种：检错码和纠错码。能发现错误但不能纠正错误的码称为检错码；不仅能发现错误而且还能纠正错误的码称为纠错码。检错码和纠错码的分类并不是绝对的，通常纠错码也可用于检错，而有的检错码也有一定的纠错能力。

检错码和纠错码可应用于不同的差错控制系统中。差错控制系统大致可分为前向纠错、重传反馈和混合纠错三种方式。

（1）前向纠错（FEC）方式

FEC（Forward Error Control）方式是，发送端发送有纠错能力的码（纠错码），接收端收到这些码后，通过纠错译码器自动地纠正传输中的错误。这种方式的优点是不需要反馈信道，

能进行一个用户对多个用户的同时通信（如广播），特别适合于移动通信；译码实时性较好，控制电路也比较简单。缺点是译码设备较复杂，编码效率较低。随着编码理论的发展和大规模集成电路的发展，译码器有可能越来越简单，成本也越来越低，因而在实际数字通信中，正在得到越来越广泛的应用。

（2）重传反馈（ARQ）方式

ARQ（Automatic Repeat Request）方式是，发送端发出能够发现错误的码（检错码），接收端译码器收到后，判断在传输中有无错误产生，并通过反馈信道把检测结果告诉发送端。发送端把接收端认为有错的消息再次传送，直到接收端认为正确接收为止。

应用 ARQ 方式必须有一条从收端至发端的反馈信道。并要求信源产生信息的速率可以进行控制，收、发两端必须互相配合，其控制电路比较复杂，传输信息的连贯性和实时性也较差。该方式的优点是译码设备简单，在多余度一定的情况下，码的检错能力比纠错能力要高得多，因而整个系统能获得极低的误码率。

（3）混合纠错（HEC）方式

HEC（Hybrid Error Control）方式是上述两种方式的结合。发送端发送的码既能检错、又有一定的纠错能力。接收端译码时若发现错误位数在码的纠错能力以内，则自动进行纠错；若错误位数超过了码的纠错能力，但能检测出来，则通过反馈信道告知发方重发。这种方式在一定程度上避免了 FEC 方式译码设备复杂和 ARQ 方式信息连贯性差的缺点，因此近年来得到了较为广泛的应用。

值得指出，在设计差错控制系统时，选择何种实现方式，应综合考虑各方面的因素，主要有：

（1）满足用户对误码率的要求；

（2）有尽可能高的信息传输速率；

（3）有尽可能简单的编译码算法且易于实现；

（4）可接受的成本。

8.1.4 纠错码的分类

现在人们已发现的纠、检错码有许多种，这些码都是给待传的信息位加上一定的冗余度（校验位），信息位与校验位共同构成码字。

按信息位与校验位的关系，可将编码分成以下两大类：

（1）线性码——校验位与信息位呈线性关系（即可用一次方程描述）。

（2）非线性码——校验位与信息位不呈线性关系。

按信息位与校验位之间的约束关系，可分为：

（1）分组码——将信息符号分成 k 位一组，每组增加 r 位校验位，这 r 位校验位仅与本组的 k 位信息位有关，与其他的信息位无关。

① 循环码——除全零码字外，其余码字都可由另一码字的码符循环移位得到；

② 非循环码——某个码字的循环移位不一定还是该码的码字。

（2）卷积码——将信息符号分成 k 位一组，每组增加 r 位校验位，这 r 位校验位不仅与本组的 k 位信息位有关，还与前面 m 组的信息位有关。

将信源的输出序列分成长为 k 的段 $\boldsymbol{u}=u_{k-1},\cdots,u_1,u_0$，序列中的每一分量都是一随机变量。

信道符号集$\{b_1,b_2,\cdots,b_D\}$，为了能够纠错，信道编码器（纠错码编码器）按一定的规则将 $\boldsymbol{u}=u_{k-1},\cdots,u_1,u_0$ 编为长为 n 的码字（码符号序列）$\boldsymbol{c}=c_{n-1},\cdots,c_1,c_0$（$n>k$），如图 8-3 所示。序列中的每一分量 $c_j\in\{b_1,b_2,\cdots,b_D\}$，$j=0,1,2,\cdots,n-1$，码字共有 n 位，其中 k 位为信息位，$n-k$ 位为校验位，假设共有 M 个消息序列，则对应的 M 个码字的集合$\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M\}$称为一个（$n$，$k$）分组码，记为 $\boldsymbol{C}$。

在上述分组码中，若 $\boldsymbol{u}$ 与 $\boldsymbol{c}$ 的对应关系是线性的，则称为线性分组码。

8.2 线性分组码的编码

8.2.1 生成矩阵、校验矩阵

目前，绝大多数的数字计算机和数字通信系统中广泛采用二进制形式的码，因此以下对线性分组码的讨论都在有限域 GF(2)上进行。域中元素为{0，1}，域中运算为模 2 加法和模 2 乘法。

信源所给出的二元信息序列首先分成等长的各个信息组，每组的信息位长度为 k，记为

$$\boldsymbol{u}=(u_{k-1}u_{k-2}\cdots u_0)$$

显然，信息组 m 可能有 2^k 种取值。编码器按一定规则，将输入的信息组编制成长为 n 的码字，记为

$$\boldsymbol{c}=(c_{n-1}c_{n-2}\cdots c_0)$$

在 $\boldsymbol{c}$ 的 n 个元素中，有 k 位是信息元，$n-k$ 位是校验元，如果各校验元与 k 位信息元之间的关系是线性的，则称这样的码为线性分组码。

例 8.1 给出了一个（7，4）线性分组码的例子。

【例 8.1】 线性分组码：$n=7$，$k=4$，记为（7，4）码。

信源符号 4 位一组：$\boldsymbol{u}$=（u_3，u_2，u_2，u_0），$u_i\in\{0,1\}$，$i=0,1,2,3$；

码符号 7 位一组：$\boldsymbol{C}$=（c_6，c_5，c_4，c_3，c_2，c_1，c_0），$c_j\in\{0,1\}$，$j=0,1,2,3,4,5,6$；

码符号与信源符号的关系为

$$\begin{cases} c_6=u_3 \\ c_5=u_2 \\ c_4=u_1 \\ c_3=u_1+u_2+u_3 \\ c_2=u_0 \\ c_1=u_0+u_2+u_3 \\ c_0=u_0+u_1+u_3 \end{cases} \tag{8-1}$$

其中，c_6，c_5，c_4，c_2 为信息位，c_3，c_1，c_0 为校验位，码符号位是信息位的线性组合，用矩阵表示式（8-1），得

$$[c_6\ \ c_5\ \ c_4\ \ c_3\ \ c_2\ \ c_1\ \ c_0]=[u_3\ \ u_2\ \ u_1\ \ u_0]\begin{bmatrix}1&0&0&1&0&1&1\\0&1&0&1&0&1&0\\0&0&1&1&0&0&1\\0&0&0&0&1&1&1\end{bmatrix} \tag{8-2}$$

将式（8-2）简记为 $\boldsymbol{c}=\boldsymbol{uG}$，称 $\boldsymbol{G}=\begin{bmatrix}1&0&0&1&0&1&1\\0&1&0&1&0&1&0\\0&0&1&1&0&0&1\\0&0&0&0&1&1&1\end{bmatrix}$ 为生成矩阵。

表 8-1 列出了按式（8-2）生成的 16 个码字。

表 8-1　（7，4）线性分组码

信　息　组	码　　字
0 0 0 0	0000000
0 0 0 1	0000111
0 0 1 0	0011001
0 0 1 1	0011110
0 1 0 0	0101010
0 1 0 1	0101101
0 1 1 0	0110011
0 1 1 1	0110100
1 0 0 0	1001011
1 0 0 1	1001100
1 0 1 0	1010010
1 0 1 1	1010101
1 1 0 0	1100001
1 1 0 1	1100110
1 1 1 0	1111000
1 1 1 1	1111111

从线性空间的概念来看，码字长度为 $n=7$，所有 2^7=128 个 7 重矢量组成一个 7 维线性空间 $\boldsymbol{V}_7$，而信息位 $k=4$，可生成 2^4=16 个码字，将每个码字都看成一个 7 位矢量（也称为码矢），这 16 个码字是从 7 位矢量空间的 128 个矢量中按照编码规则选出来的。因此可以说，2^4 个码字是线性空间 $\boldsymbol{V}_7$ 中的一个 4 维子空间 $\boldsymbol{V}_7^4$。而生成矩阵 $\boldsymbol{G}$ 由 k=4 维子空间 $\boldsymbol{V}_7^4$ 中的 4 个线性无关行矢组成。

定义矩阵

$$\boldsymbol{H}=\begin{bmatrix}1&0&1&0&1&0&1\\1&1&0&0&1&1&0\\1&1&1&1&0&0&0\end{bmatrix}$$

可以验算，矩阵 $\boldsymbol{H}$ 与 $\boldsymbol{G}$ 正交，即满足 $\boldsymbol{HG}^T=0$ 或 $\boldsymbol{GH}^T=0$，因此有

$$\boldsymbol{Hc}^T=\boldsymbol{H}(\boldsymbol{uG})^T=\boldsymbol{HG}^T\boldsymbol{u}^T=\boldsymbol{0}\boldsymbol{u}^T=\boldsymbol{0} \tag{8-3}$$

称 $\boldsymbol{H}$ 为校验矩阵，$\boldsymbol{H}$ 由 $n=7$ 维空间的 $n-k$=7−4=3 维子空间 $\boldsymbol{V}_7^3$ 中的 3 个线性无关行矢组成。

结合图 8-3 就此例可将编、解码步骤归纳如下：

（1）信源输出序列 $\boldsymbol{u}$=（u_3，u_2，u_2，u_0）；

（2）纠错编码器将 $\boldsymbol{u}$ 编码为 $\boldsymbol{c}=\boldsymbol{u}\,\boldsymbol{G}$ =（c_6，c_5，c_4，c_3，c_2，c_1，c_0）；

（3）将 $\boldsymbol{c}$ 遣入信道传输，由于干扰，输出 $\boldsymbol{y}$=（y_6，y_5，y_4，y_3，y_2，y_1，y_0）；

（4）计算 $\boldsymbol{Hy}^T=\boldsymbol{s}$，称 $\boldsymbol{s}$ 为 $\boldsymbol{y}$ 的**伴随式**，若 $\boldsymbol{s}=\boldsymbol{0}$，根据式（8-3），知道 $\boldsymbol{y}$ 是选用码矢，传输无误；若 $\boldsymbol{s}\neq\boldsymbol{0}$，则 $\boldsymbol{y}$ 不是选用码矢，说明在传输过程中发生了误码；

（5）记 $\boldsymbol{e}$ 为错误图案，则

$$\boldsymbol{y}=\boldsymbol{c}+\boldsymbol{e}$$

$$\boldsymbol{s}=\boldsymbol{Hy}^T=\boldsymbol{Hc}^T+\boldsymbol{He}^T=\boldsymbol{He}^T \tag{8-4}$$

式（8-4）说明伴随式仅与错误图案有关，假设误码只有一位，则列矢 $\boldsymbol{e}$ 中只有一位为 1，由式（8-4）可看出，观察列矢 $\boldsymbol{s}$ 与校验矩阵 $\boldsymbol{H}$ 中的哪一列相同，就可确定 $\boldsymbol{H}$ 中的哪一位出错。

例如，信息位 $\boldsymbol{u}=0011$，编码为 $\boldsymbol{c}=0011110$，假设传输中第 2 位出错，接收矢量为 $\boldsymbol{y}=0111110$，可算出伴随式

$$\boldsymbol{s}=\boldsymbol{Hy}^T=\begin{bmatrix}1&0&1&0&1&0&1\\1&1&0&0&1&1&0\\1&1&1&1&0&0&0\end{bmatrix}\begin{bmatrix}0\\1\\1\\1\\1\\1\\0\end{bmatrix}=\begin{bmatrix}0\\1\\1\end{bmatrix}$$

伴随式 $\boldsymbol{s}$ 与 $\boldsymbol{H}$ 矩阵中的第 2 列相同，由此判断接收矢量 $\boldsymbol{y}$ 中第 2 位出错。

从此例可看出，上述（7，4）线性分组码可纠正一位错码，适用于固有误码率较小的场合。

可见纠错编码就是给信息位加上一定的冗余度，结果是降低了传输效率，但提高了传输可靠性。

将每一个码字看成一个 n 重矢量，则全部 n 重矢量对应 n 维线性空间 $\boldsymbol{V}_n$，线性分组码则对应 n 维线性空间的 k（$k<n$）维子空间 $\boldsymbol{V}_n^k$。

$r=k/n$ 称为分组码的码率，它是线性分组码最重要的参数之一。它说明了在一个码字中信息位所占的比重。r 越大，表明信息位所占的比重越大，码的传输信息的有效性越高。

定义 8.1　2^k 个 n 重矢量的集合 $\boldsymbol{C}$ 称为线性分组码，当且仅当它是 n 维线性空间 $\boldsymbol{V}_n$ 中的一个 k 维子空间 $\boldsymbol{V}_n^k$。

在 k 维子空间中，只能找到 k 个 n 重线性无关的矢量，定为基矢，其余 n 重矢量都可以表示为基矢的线性组合。

（n，k）线性分组码的编码问题，就是如何从 n 维线性空间 $\boldsymbol{V}_n$ 中，找到满足一定要求的，由 2^k 个矢量组成的 k 维线性子空间 $\boldsymbol{V}_n^k$；或者说在满足一定条件下，如何根据已知的 k 个信息元求得 $n-k$ 个校验元。

由于（n，k）线性分组码的 2^k 个码字组成了 n 维线性空间 $\boldsymbol{V}_n$ 的一个 k 维子空间 $\boldsymbol{V}_n^k$，因此这 2^k 个码字完全可由 k 个线性无关的矢量所组成的基底所张成。

记线性码 $\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M\}$，$\boldsymbol{C}$ 具有如下性质。

（1）封闭性：$\boldsymbol{c}_1\in\boldsymbol{C}$，$\boldsymbol{c}_2\in\boldsymbol{C}$，则 $\boldsymbol{c}_1+\boldsymbol{c}_2\in\boldsymbol{C}$；

（2）结合律：$\boldsymbol{c}_1$，$\boldsymbol{c}_2$，$\boldsymbol{c}_3\in\boldsymbol{C}$，则$(\boldsymbol{c}_1+\boldsymbol{c}_2)+\boldsymbol{c}_3=\boldsymbol{c}_1+(\boldsymbol{c}_2+\boldsymbol{c}_3)$；

（3）存在全零码字 $\boldsymbol{0}$，使 $\boldsymbol{c}+\boldsymbol{0}=\boldsymbol{c}$；

（4）存在逆码 $\boldsymbol{c}^{-1}$，使 $\boldsymbol{c}+\boldsymbol{c}^{-1}=\boldsymbol{0}$；

（5）交换律：$\boldsymbol{c}_1+\boldsymbol{c}_2=\boldsymbol{c}_2+\boldsymbol{c}_1$。

将线性码的这 5 条性质与交换群的性质相比较，可看出，码 $\boldsymbol{C}$ 是一个加法交换群。

注意：由于（n，k）线性分组码是 n 维线性空间的一个 k 维子空间，故其运算满足封闭性条件。

通过例 8.1，我们对生成矩阵和校验矩阵有了初步了解。

定义 8.2 （n，k）线性分组码 $\boldsymbol{C}$ 中一组基底所构成的 $k\times n$ 阶矩阵称为码 $\boldsymbol{C}$ 的生成矩阵，用 $\boldsymbol{G}$ 表示。

设此基底为 $\boldsymbol{g}_0$，$\boldsymbol{g}_1$，…，$\boldsymbol{g}_{k-1}$，

$$\begin{aligned}\boldsymbol{g}_0&=(g_{0,n-1},\ g_{0,n-2},\ \cdots,\ g_{0,0})\\ \boldsymbol{g}_1&=(g_{1,n-1},\ g_{1,n-2},\ \cdots,\ g_{1.0})\\ &\vdots\\ \boldsymbol{g}_{k-1}&=(g_{k-1,n-1},\ g_{k-1,n-2},\ \cdots,\ g_{k-1,0})\end{aligned}$$

写成矩阵形式，为

$$\boldsymbol{G}=\begin{bmatrix}\boldsymbol{g}_0\\ \boldsymbol{g}_1\\ \vdots\\ \boldsymbol{g}_{k-1}\end{bmatrix}=\begin{bmatrix}g_{0,n-1} & g_{0,n-2} & \cdots & g_{0,0}\\ g_{1,n-1} & g_{1,n-2} & \cdots & g_{1,0}\\ \vdots & \vdots & & \vdots\\ g_{k-1,n-1} & g_{k-1,n-2} & \cdots & g_{k-1,0}\end{bmatrix} \tag{8-5}$$

对于任一给定的信息序列，记为 k 重行矢 $\boldsymbol{u}=(u_{k\text{-}1}, u_{k\text{-}2}, \cdots, u_1, u_0)$，则 n 重行矢

$$\begin{aligned}\boldsymbol{c}&=\boldsymbol{u}\boldsymbol{G}\\ &=\begin{pmatrix}u_{k-1} & u_{k-2} & \cdots & u_0\end{pmatrix}\begin{bmatrix}g_{0,n-1} & g_{0,n-2} & \cdots & g_{0,0}\\ g_{1,n-1} & g_{1,n-2} & \cdots & g_{1,0}\\ \vdots & \vdots & & \vdots\\ g_{k-1,n-1} & g_{k-1,n-2} & \cdots & g_{k-1,0}\end{bmatrix}\\ &=\left(\sum_{i=0}^{k-1}u_{k-1-i}g_{i,n-1}\quad \sum_{i=0}^{k-1}u_{k-1-i}g_{i,n-2}\quad \cdots\quad \sum_{i=0}^{k-1}u_{k-1-i}g_{i,0}\right)\\ &=\begin{pmatrix}c_{n-1} & c_{n-2} & \cdots & c_0\end{pmatrix}\end{aligned} \tag{8-6}$$

就是信息序列 $\boldsymbol{u}=(u_{k-1}, u_{k-2}, \cdots, u_0)$对应的码字，其中加法和乘法运算遵照 GF(2)中的运算规则。信息序列取不同值就得到不同的码字，当信息序列 $\boldsymbol{u}=(u_{k-1}, u_{k-2}, \cdots, u_1, u_0)$遍取所有值分别代入上式时，得到的码字集合记为码集 $\boldsymbol{C}$。

从式（8-6）可看出，基底{ $\boldsymbol{g}_0$，$\boldsymbol{g}_1$，…，$\boldsymbol{g}_{k-1}$}实际上就是码集 $\boldsymbol{C}$ 中的 k 个码字，分别对应信息序列 $\boldsymbol{u}_{k-1}=(1, 0, \cdots, 0, 0)$，$\boldsymbol{u}_{k-2}=(0, 1, \cdots, 0, 0)$，…，$\boldsymbol{u}_1=(0, 0, \cdots, 0, 1)$。

一个线性空间的基底可以不只一组，因此作为码的生成矩阵 $\boldsymbol{G}$，也可以不止一种形式。可以从码集 $\boldsymbol{C}$ 中另取 k 个线性无关的码矢作为基底，它们也生成相同的线性空间，即生成同一个（n，k）线性分组码。

由例 8.1 可看出，基底取的是信息组（1000），（0100），（0010），（0001）所对应的码字（1001011），（0101010），（0011001），（0000111），如果在表 8-1 中另取 4 个线性无关的码字做为基底：（0011110），（0101101），（1100001），（1111000），构成生成矩阵 $\boldsymbol{G}=\begin{bmatrix}0&0&1&1&1&1&0\\ 0&1&0&1&1&0&1\\ 1&1&0&0&0&0&1\\ 1&1&1&1&0&0&0\end{bmatrix}$，则根据式（8-6）可生成表 8-2 中的码字，表中的所有码字与表 8-1

相同，只不过每个码字对应的信息序列不同而已。

表 8-2　另取基底的（7，4）线性分组码

信　息　组	码　　字
0 0 0 0	0000000
0 0 0 1	1111000
0 0 1 0	1100001
0 0 1 1	0011001
0 1 0 0	0101101
0 1 0 1	1010101
0 1 1 0	1001100
0 1 1 1	0110100
1 0 0 0	0011110
1 0 0 1	1100110
1 0 1 0	1111111
1 0 1 1	0000111
1 1 0 0	0110011
1 1 0 1	1001011
1 1 1 0	1010010
1 1 1 1	0101010

综上所述，对于（n，k）线性分组码 $\boldsymbol{C}$，全部 2^n 个 n 重矢量对应 n 重线性空间 $\boldsymbol{V}_n$。而码集 $\boldsymbol{C}$ 只包括其中的 2^k 个 n 重矢量，这 2^k 个码矢对应 k 维子空间 $\boldsymbol{V}_n^k$，由 k 个线性无关的矢量所张成，这 k 个矢量写成矩阵形式就是（n，k）线性分组码的生成矩阵 $\boldsymbol{G}$，这是一个 $k\times n$ 矩阵。

在 $\boldsymbol{V}_n$ 线性空间中，一定存在着与 $\boldsymbol{V}_n^k$ 对应的 $n-k$ 维的零化空间 $\boldsymbol{V}_n^{n-k}$，$\boldsymbol{V}_n^{n-k}$ 由 $n-k$ 个线性无关的矢量所张成，这 $n-k$ 个矢量写成矩阵形式 $\boldsymbol{H}$，这是一个$(n-k)\times n$ 矩阵。

生成矩阵 $\boldsymbol{G}$ 和矩阵 $\boldsymbol{H}$ 满足下式：

$$\boldsymbol{G}\boldsymbol{H}^T=\boldsymbol{0} \quad \text{或} \quad \boldsymbol{H}\boldsymbol{G}^T=\boldsymbol{0} \tag{8-7}$$

式（8-7）说明 $\boldsymbol{G}$ 的行矢与 $\boldsymbol{H}$ 的列矢正交，利用式（8-6）可得

$$\boldsymbol{H}\boldsymbol{c}^T=\boldsymbol{H}(\boldsymbol{u}\boldsymbol{G})^T=\boldsymbol{H}\boldsymbol{G}^T\boldsymbol{u}^T=\boldsymbol{0} \tag{8-8}$$

式（8-8）说明矩阵 $\boldsymbol{H}$ 可以用来校验一个 n 重矢量是否是被选用码矢，所以称 $\boldsymbol{H}$ 为校验矩阵。

8.2.2　系统码

一个子空间的基底选用并不是唯一的，所以对应的生成矩阵 $\boldsymbol{G}$ 也不是唯一的。

【例 8.2】　二元（6，3）线性分组码，下面给出的 $\boldsymbol{G}_1$ 和 $\boldsymbol{G}_2$ 都可以作为它的生成矩阵。

$$\boldsymbol{G}_1=\begin{bmatrix}1&0&1&0&1&1\\1&1&0&1&0&1\\1&1&1&0&0&0\end{bmatrix} \qquad \boldsymbol{G}_2=\begin{bmatrix}1&0&0&1&1&0\\0&1&0&0&1&1\\0&0&1&1&0&1\end{bmatrix}$$

表 8-3 给出了分别由 $\boldsymbol{G}_1$ 和 $\boldsymbol{G}_2$ 所生成的线性码。

表 8-3　由不同生成矩阵生成的线性分组码

信　息　组	由 $\boldsymbol{G}_1$ 生成的（6，3）码	由 $\boldsymbol{G}_2$ 生成的（6，3）码
000	000000	000000
001	111000	001101
010	110101	010011
011	001101	011110
100	101011	100110
101	010011	101011
110	011110	110101
111	100110	111000

从表 8-3 可以看出，由 $\boldsymbol{G}_1$ 和 $\boldsymbol{G}_2$ 所生成的线性码对应同一个 3 维子空间 $\boldsymbol{V}_6^3$，只不过信息组与码矢之间有着不同的对应关系。

观察由 $\boldsymbol{G}_2$ 所生成的线性码，由

$$\begin{aligned}\boldsymbol{c} &= \boldsymbol{u}\boldsymbol{G}\\ &= \begin{pmatrix}u_2 & u_1 & u_0\end{pmatrix}\begin{bmatrix}1&0&0&1&1&0\\0&1&0&0&1&1\\0&0&1&1&0&1\end{bmatrix}\\ &=（u_2，\ u_1，\ u_0，\ u_2+u_0，\ u_2+u_1，\ u_1+u_0）\\ &=（c_5，\ c_4，\ c_3，\ c_2，\ c_1，\ c_0）\end{aligned}$$

得

$$\begin{cases}c_5 = u_2\\ c_4 = u_1\\ c_3 = u_0\\ c_2 = u_2 + u_0\\ c_1 = u_2 + u_1\\ c_0 = u_1 + u_0\end{cases}$$

码矢的前 3 位是信息位，后 3 位是校验位，将这样的码称为系统码。

在一般情况下，生成矩阵 $\boldsymbol{G}$ 是 $k\times n$ 矩阵，可通过初等变换将 $\boldsymbol{G}$ 变成如下形式。

$$\boldsymbol{G} = \begin{bmatrix}1&0&\cdots&0&p_{11}&p_{12}&\cdots&p_{1(n-k)}\\0&1&\cdots&0&p_{21}&p_{22}&\cdots&p_{2(n-k)}\\\vdots&\vdots&\cdots&\vdots&\vdots&\vdots&\cdots&\vdots\\0&0&\cdots&1&p_{k1}&p_{k2}&\cdots&p_{k(n-k)}\end{bmatrix} = \left[\boldsymbol{I}_k \vdots \boldsymbol{P}\right]$$

$\boldsymbol{I}_k$ 是 $k\times k$ 单位方阵，$\boldsymbol{P}$ 是 $k\times(n-k)$ 矩阵（$n<k$），称这种形式的 $\boldsymbol{G}$ 为**标准生成矩阵**，因为初等变换不改变矩阵的秩，因此 $\left[\boldsymbol{I}_k \vdots \boldsymbol{P}\right]$ 仍由 k 个线性无关的行矢组成。这样生成的码 $\boldsymbol{c} = \boldsymbol{u}\boldsymbol{G} = \boldsymbol{u}[\boldsymbol{I}_k \boldsymbol{P}] = (c_{n-1}, c_{n-2}, \cdots, c_1, c_0)$，前面 k 位 $c_{n-1}, c_{n-2}, \cdots, c_{n-k}$ 是信息位，后面 $n-k$ 位 $c_{n-k-1}, c_{n-k-2}, \cdots, c_1, c_0$ 是校验位，称这种码为线性系统分组码，简称**系统码**。

这时校验矩阵相应地变成 $\boldsymbol{H} = \left[-\boldsymbol{P}^T \vdots \boldsymbol{I}_{n-k}\right]$，其中 $\boldsymbol{I}_{n-k}$ 是 $(n-k)\times(n-k)$ 单位方阵，$-\boldsymbol{P}^T$ 是矩阵 $\boldsymbol{P}$ 的逆元转置矩阵，对于模 2 加运算，0 的逆原为 1，1 的逆元为 0，故有 $-\boldsymbol{P}^T = \boldsymbol{P}^T$，$\boldsymbol{H} = \left[\boldsymbol{P}^T \vdots \boldsymbol{I}_{n-k}\right]$。

可以验证

$$\boldsymbol{G}\boldsymbol{H}^{\mathrm{T}}=\left[\boldsymbol{I}_k \vdots \boldsymbol{P}\right]\left[\boldsymbol{P}^{T} \vdots \boldsymbol{I}_{n-k}\right]=\boldsymbol{0}$$

称 $\boldsymbol{H}$ 为一致校验矩阵，简称校验矩阵。

常用的系统码有两种形式：信息组被排在码字的最左边 k 位，或信息组被排在码字的最右边 k 位。本书中的系统码均采用第一种形式。

一般来说，系统码的译码相对非系统码要简单一些，但两者的纠错能力完全等价，因此一般总希望线性分组码采用系统码形式。

【例 8.3】 对于例 8.1 给出的（7，4）线性分组码，

生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&1&0&1&1\\0&1&0&1&0&1&0\\0&0&1&1&0&0&1\\0&0&0&0&1&1&1\end{bmatrix}\overset{\text{初等变换}}{\Rightarrow}\begin{bmatrix}1&0&0&0&1&1&1\\0&1&0&0&1&1&0\\0&0&1&0&1&0&1\\0&0&0&1&0&1&1\end{bmatrix}=\left[\boldsymbol{I}_k \vdots \boldsymbol{P}\right]$

校验矩阵 $\boldsymbol{H}=\left[\boldsymbol{P}^{T} \vdots \boldsymbol{I}_{n-k}\right]=\begin{bmatrix}1&1&1&0&1&0&0\\1&1&0&1&0&1&0\\1&0&1&1&0&0&1\end{bmatrix}$

生成的码字如表 8-4 所示。

表 8-4 （7，4）系统码

信 息 组	码 字
0 0 0 0	0000000
0 0 0 1	0001011
0 0 1 0	0010101
0 0 1 1	0011110
0 1 0 0	0100110
0 1 0 1	0101101
0 1 1 0	0110011
0 1 1 1	0111000
1 0 0 0	1000111
1 0 0 1	1001100
1 0 1 0	1010010
1 0 1 1	1011001
1 1 0 0	1100001
1 1 0 1	1101010
1 1 1 0	1110100
1 1 1 1	1111111

表中的码字前 4 位是信息位，后 3 位是校验位。

8.2.3 对偶码

设原码有 k 位信息位，其生成矩阵为 $\boldsymbol{G}$，校验矩阵为 $\boldsymbol{H}$，对应线性码 $\boldsymbol{C}$。

若用 $\boldsymbol{H}$ 作为生成矩阵，生成另一码 $\boldsymbol{C}^{\perp}$，则对应的校验矩阵为 $\boldsymbol{G}$，称 $\boldsymbol{C}^{\perp}$ 为 $\boldsymbol{C}$ 的**对偶码**，$\boldsymbol{C}^{\perp}$ 有

$n-k$ 位信息位，k 位校验位。

因为 $\boldsymbol{c}^{\perp}=\boldsymbol{uH}$，$\boldsymbol{c}=\boldsymbol{uG}$，$\boldsymbol{c}\ (\boldsymbol{c}^{\perp})^{\mathrm{T}}=\boldsymbol{uG}(\boldsymbol{uH})^{\mathrm{T}}=\boldsymbol{u}(\boldsymbol{GH}^{\mathrm{T}})\ \boldsymbol{u}^{\mathrm{T}}=0$，说明互为对偶的码矢内积为 0，两码矢正交。

【例 8.4】 给定生成矩阵 $\boldsymbol{G}_{(7.3)}=\begin{bmatrix}1&0&0&1&1&1&0\\0&1&0&0&1&1&1\\0&0&1&1&1&0&1\end{bmatrix}$，求由 $\boldsymbol{G}_{(7.3)}$ 生成的原码 $\boldsymbol{C}$ 和它的对偶码 $\boldsymbol{C}^{\perp}$。

由 $\boldsymbol{G}_{(7.3)}$ 生成（7，3）线性分组码：

$$\boldsymbol{c}_{(7.3)}=\boldsymbol{u}\cdot\boldsymbol{G}_{(7.3)}=\begin{bmatrix}u_2&u_1&u_0\end{bmatrix}\cdot\begin{bmatrix}1&0&0&1&1&1&0\\0&1&0&0&1&1&1\\0&0&1&1&1&0&1\end{bmatrix}$$

$$=\begin{bmatrix}u_2,&u_1,&u_0,&u_2\oplus u_0,&u_2\oplus u_1\oplus u_0,&u_2\oplus u_1,&u_1\oplus u_0\end{bmatrix}$$

具体码见表 8-5。

而 $\boldsymbol{G}_{(7.3)}$ 就是其对偶码 $\boldsymbol{C}_{(7.4)}^{\perp}$ 的校验矩阵 $\boldsymbol{H}_{(7.4)}^{\perp}$，先将 $\boldsymbol{G}_{(7.3)}$ 通过初等变换化为校验矩阵 $\boldsymbol{H}_{(7.4)}^{\perp}$ 的标准形式：

$$\boldsymbol{G}_{(7.3)}\overset{\text{初等变换}}{\Longrightarrow}\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{bmatrix}=\begin{bmatrix}\boldsymbol{P}^{T}\vdots\boldsymbol{I}_3\end{bmatrix}=\boldsymbol{H}_{(7,4)}^{\perp}$$

根据 $\boldsymbol{H}_{(7,4)}^{\perp}$ 可得对偶码的生成矩阵 $\boldsymbol{G}_{(7,4)}^{\perp}$：

$$\boldsymbol{G}_{(7.4)}^{\perp}=\begin{bmatrix}\boldsymbol{I}_4\vdots\boldsymbol{P}^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

由 $\boldsymbol{G}_{(7.4)}^{\perp}$ 生成对偶码：

$$\boldsymbol{C}_{(7.3)}^{\perp}=\boldsymbol{C}_{(7.4)}=\boldsymbol{u}\cdot\boldsymbol{G}_{(7.4)}^{\perp}=\begin{bmatrix}u_3&u_2&u_1&u_0\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

$$=\begin{bmatrix}u_3,&u_2,&u_1,&u_0,&u_3\oplus u_2\oplus u_1,&u_2\oplus u_1\oplus u_0,&u_3\oplus u_2\oplus u_0\end{bmatrix}$$

具体码见表 8-6。

8.2.4 编码的实现

综上所述，当已知（n，k）线性分组码的生成矩阵 $\boldsymbol{G}$ 或校验矩阵 $\boldsymbol{H}$ 时，编码问题是容易实现的。由式（8-6），设码的 $\boldsymbol{G}$ 矩阵为

$$\boldsymbol{G}=\begin{bmatrix}&p_{1,n-k-1}&p_{1,n-k-2}&\cdots&p_{1,0}\\\boldsymbol{I}_k\vdots&p_{2,n-k-1}&p_{2,n-k-2}&\cdots&p_{2,0}\\&\vdots&\vdots&&\vdots\\&p_{k,n-k-1}&p_{k,n-k-2}&\cdots&p_{k,0}\end{bmatrix}$$

表 8-5　由 $\boldsymbol{G}_{(7,3)}$生成（7，3）码

信息位 u_2, u_1, u_0	码矢 $\boldsymbol{C}$
000	0000000
001	0011101
010	0100111
011	0111010
100	1001110
101	1010011
110	1101001
111	1110100

表 8-6　由 $\boldsymbol{G}_{(7,4)}^{\perp}$ 生成（7，4）码

信息位 u_3, u_2, u_1, u_0	码矢 $\boldsymbol{C}$
0000	0000000
0001	0001011
0010	0010110
0011	0011101
0100	0100111
0101	0101100
0110	0110001
0111	0111010
1000	1000101
1001	1001110
1010	1010011
1011	1011000
1100	1100010
1101	1101001
1110	1110100
1111	1111111

当信息组 $\boldsymbol{u}=(u_{n-1}u_{n-2}\cdots u_{n-k})$ 时，相应的码字 $\boldsymbol{c}$ 是

$$\boldsymbol{c}=\boldsymbol{u}\cdot\boldsymbol{G}=(c_{n-1}\,c_{n-2}\cdots\ c_1\,c_0)$$

其中，
$$c_j=u_j,\quad n-k\leqslant j\leqslant n-1$$

及
$$c_j=u_{n-1}\,p_{1,j}+u_{n-2}\,p_{2,j}+\cdots+u_{n-k}\,p_{k,j}\qquad 0\leqslant j<n-k$$

编码实现电路如图 8-4 所示。电路由移位寄存器、模二加法器和模二乘法器组成。在图中，“→□→”表示移位寄存器单元，“⊕”表示模二加法器，“→○→”及近旁的 $\boldsymbol{P}$ 矩阵元素 p_{ij} 表示模二乘法器，对于二元域，$p_{ij}=1$ 表示该处连通，$p_{ij}=0$ 表示该处断开。

根据图 8-4 的电路，可画出例 8.3 给出的（7，3）线性分组码 $\boldsymbol{G}_{(7,3)}$ 的编码器电路，如图 8-5 所示。

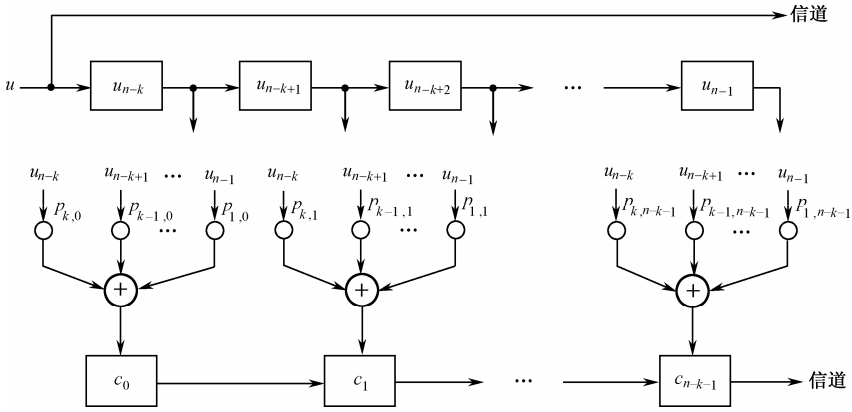


图 8-4　(n, k)线性分组码编码电路

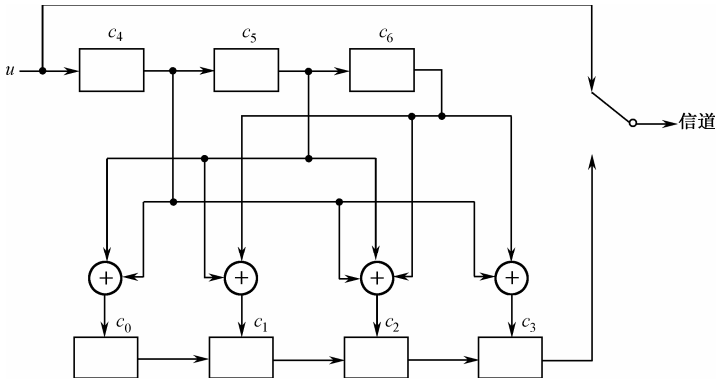


图 8-5　(7, 3)线性分组码编码电路

8.3　线性码的纠检错能力

8.3.1　码的距离和重量

为了考察码的构成与它的纠错能力之间的关系，引入码的两个重要参数：码的距离和重量。

定义 8.3　两个码字之间，对应位取值不同的个数称为它们之间的汉明距离，简称距离，用 $d(\boldsymbol{c}_1, \boldsymbol{c}_2)$ 表示。

如已知（7，3）线性分组码的两个码字

$$\boldsymbol{c}_1=(1110100),\ \boldsymbol{c}_2=(0101010)$$

则 $d(\boldsymbol{c}_1, \boldsymbol{c}_2)=5$。

定义 8.4　码字中非零码元的个数称为该码字的汉明重量，简称重量，用 $w(\boldsymbol{c})$表示。

如上例中，$w(\boldsymbol{c}_1)=4$，$w(\boldsymbol{c}_2)=3$。

在（n，k）线性分组码 $\boldsymbol{C}$中，任两个码字之间都有一定的距离，但必有一个是最小的。

定义 8.5　码 $\boldsymbol{C}$的最小距离 $d_{\min}$ 定义为

$$d_{\min}=\min\left\{d(\boldsymbol{c}_i,\boldsymbol{c}_j),\ i\neq j,\ \boldsymbol{c}_i,\boldsymbol{c}_j\in\boldsymbol{C}\right\} \tag{8-9}$$

为方便起见，将最小距离 $d_{\min}$ 简记为 d。

如表 8-1 中（7，4）码的最小距离 $d=3$。

定理 8.1　线性分组码的最小距离等于其非零码字的最小重量。

证明：设码矢 $\boldsymbol{c}_1$，$\boldsymbol{c}_2\in\boldsymbol{C}$，求 $\boldsymbol{c}_1$，$\boldsymbol{c}_2$ 间的汉明距离 $d(\boldsymbol{c}_1, \boldsymbol{c}_2)$ 就等效于求 $\boldsymbol{c}_1+\boldsymbol{c}_2=\boldsymbol{c}_3$ 的汉明重量 $w(\boldsymbol{c}_3)$，由于码 $\boldsymbol{C}$是一个加法交换群，其运算满足封闭性条件，所以有 $\boldsymbol{c}_3\in\boldsymbol{C}$，因此码的最小距离必等于其非零码字的最小重量。　　证毕

根据定理 8.1，要得到码的最小距离，只要检查 2^k-1 个非零码字的重量即可。例如，表 8-1 中（7，4）码的最小距离 $d=3$，就是从表 8-1 中检查 15 个非零码字的重量后得到的。

码的距离和重量满足以下不等式：

$$d(\boldsymbol{c}_1,\ \boldsymbol{c}_2)\leqslant d(\boldsymbol{c}_1,\ \boldsymbol{c}_3)+d(\boldsymbol{c}_3,\ \boldsymbol{c}_2) \tag{8-10}$$

$$w(\boldsymbol{c}_1+\boldsymbol{c}_2)\leqslant w(\boldsymbol{c}_1)+w(\boldsymbol{c}_2) \tag{8-11}$$

这些性质对于研究线性分组码是有用的。

事实上，两个码字之间的距离表示了它们之间差别的大小。距离愈大，两个码字之间的差别就愈大，则传送时从一个码字变成另一个码字的可能性就愈小。因此，一个线性分组码的最小距离是衡量码抗干扰能力的重要参数。码的最小距离 d 愈大，其抗干扰能力愈强。8.3.2 节将详细介绍码的纠错能力与最小距离 d 之间的定量关系。

8.3.2 线性码的纠错、检错能力

线性码的纠错、检错能力与译码规则有关，第 5 章介绍了最大后验概率译码准则，即最小错误概率译码准则。下面根据这一准则来分析线性分组码的纠错、检错能力。

设码矢 $\boldsymbol{c}_m, \boldsymbol{c}_m' \in \boldsymbol{C}$，若发送的码矢为 $\boldsymbol{c}_m$，由于干扰，信道输出端接收到矢量 $\boldsymbol{y}$，如图 8-6 所示，因为矢量 $\boldsymbol{y}$ 离码矢 $\boldsymbol{c}_m$ 最近，按最小错误概率译码准则译码，可将 $\boldsymbol{y}$ 译成 $\boldsymbol{c}_m$，如果信道条件很差，$\boldsymbol{c}_m$ 错得厉害，错成 $\boldsymbol{y}'$，由于 $\boldsymbol{y}'$ 离码矢 $\boldsymbol{c}_m'$ 最近，则按最小错误概率译码准则将 $\boldsymbol{y}'$ 译成 $\boldsymbol{c}_m'$，就发生了误码。

那么错到什么程度不至于产生误码？从线性空间的角度来看，最小错误概率译码准则就是最小距离译码准则，判决规则实际上是对矢量空间的一种划分，假设码 $\boldsymbol{C}$ 能发现（检测）e 位错误，以 $\boldsymbol{c}_m$ 为球心，e 为半径作球 $Q_e(m)$，若接收矢量 $\boldsymbol{y} \in$ 球 $Q_e(m)$，译为 $\boldsymbol{c}_m$，不妨设码矢 $\boldsymbol{c}_m$ 与另一码矢 $\boldsymbol{c}_m'$ 之间具有最小汉明距离 d（这样发送码矢 $\boldsymbol{c}_m$ 才容易误判为 $\boldsymbol{c}_m'$，因为其他的码矢距离 $\boldsymbol{c}_m$ 都更远）。则一定要满足

$$d \geq e+1 \tag{8-12}$$

才不会将接收矢量 $\boldsymbol{y}$ 译码为离 $\boldsymbol{c}_m$ 最近的另一码矢与 $\boldsymbol{c}_m'$，如图 8-7 所示。

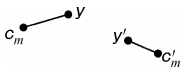


图 8-6　最小距离译码

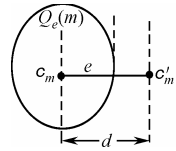


图 8-7　检错判决空间划分示意图

又假设码 $\boldsymbol{C}$ 的纠错能力为 t 位，即错 t 位就不会产生误码，设码矢 $\boldsymbol{c}_m$ 与另一码矢 $\boldsymbol{c}_m'$ 之间具有最小汉明距离 d，分别以 $\boldsymbol{c}_m$ 和 $\boldsymbol{c}_m'$ 为球心，t 为半径作球 $Q_t(m)$ 和 $Q_t(m')$，如图 8-8 所示。判决空间可以这样来划分：若接收矢量 $\boldsymbol{y} \in$ 球 $Q_t(m)$，则译为 $\boldsymbol{c}_m$；若 $\boldsymbol{y} \in$ 球 $Q_t(m')$，则译为 $\boldsymbol{c}_m'$。显然这两个球应不相交才不会产生误码（若相交，当接收矢量落在重合部分时，则可译成 $\boldsymbol{c}_m$，也可译成 $\boldsymbol{c}_m'$）。要使两球不相交，必须满足 $d_{\min} \geq 2t+1$（t, d 皆为整数），即

$$t \leq \frac{d-1}{2} \tag{8-13}$$

根据式（8-13），在码 $\boldsymbol{C}$ 中小于等于 $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ 个错误均可纠正，大于 t 个就不一定可纠。

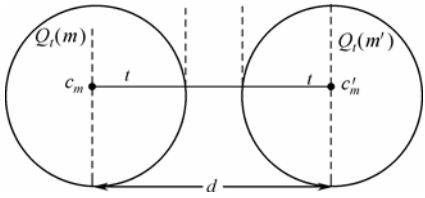


图 8-8　纠错判决空间划分示意图

如果码 $\boldsymbol{C}$ 在发现 e 位错误的同时，还能纠正 t 位错误，则应有 $d \geq t+e+1$，由此得

$$e \leq d-t-1 \tag{8-14}$$

显然，这时分别以 $\boldsymbol{c}_m$ 和 $\boldsymbol{c}_m'$ 为球心，t 和 e 为半径所作的两球是不相交的。下面用反证法证明。

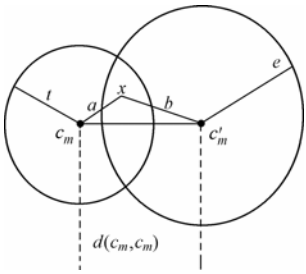


图 8-9　设两球相交

设两球相交，如图 8-9 所示，x 是相交部分中的任意一点，x 与点 $\boldsymbol{c}_m$ 和 $\boldsymbol{c}_m'$ 组成三角形，a，b 为三角形的两条边，则码矢 $\boldsymbol{c}_m$ 和 $\boldsymbol{c}_m'$ 之间的距离

$$d(\boldsymbol{c}_m,\boldsymbol{c}_m') \leq a+b < t+e \leq d-1 < d$$

即

$$d(\boldsymbol{c}_m,\boldsymbol{c}_m') < d$$

说明码矢 $\boldsymbol{c}_m$ 和 $\boldsymbol{c}_m'$ 之间的距离小于 d，这与 d 是最小汉明距离相矛盾，故两球不可能相交。

定理 8.2　对于任一个（n，k）线性分组码，若要在码字内

（1）检测 e 位错误，则要求码的最小距离 $d \geq e+1$；

（2）纠正 t 位错误，则要求码的最小距离 $d \geq 2t+1$；

（3）纠正 t 位错误的同时还能检测 e（$\geq t$）位错误，则要求 $d \geq t+e+1$。

这里所谓的“同时”，是指当错误位数小于等于 t 时，译码器能正确实施纠错；当错误位数小于等于 e 时，译码器同时能检测出错误。

表 8-1 中（7，4）码的最小距离 $d=3$，它能检测 $e \leq d-1=2$ 位错误，纠正 $t=\left\lfloor \frac{d-1}{2} \right\rfloor=\left\lfloor \frac{3-1}{2} \right\rfloor=1$ 位错误。

例 5.2 给出的（n，1）重复码，该码信息位数 $k=1$，因此只有 2 个许用码字；$r=n-1$ 个校验元的产生规则是信息元的重复，故得名为重复码。当 n 越大时，码的最小距离 d 也越大，其抗干扰能力也越强。

（2，1）重复码，许用码字为（00）和（11），显然 $d=2$。由上述定理，该码只能检测单个错误。当译码器收到禁用码字（01）或（10）时，发现其不是许用码字，故能检测出一个错误，但却不能确定收到的禁用码字是从（00）还是从（11）错一位造成的，因此无法纠正错误。

（3，1）重复码，许用码字为（000）和（111），$d=3$。当用于纠错时，它能纠正单个错误；用于检错时，可检测两个错误。该码有 $2^n-2^k=6$ 个禁用码字：（001），（010），（100），（011），（101），（110）。根据择多译码规则：当接收方接收的是（001），（010），（100）时，译码就判为（000）；当接收的是（011），（101），（110）时，译码就判为（111）。显然，用这种方法译码，能正确纠正一个错误。另外，当许用码字（000）或（111）传送出现两个错时，译码器仍能发现有错，但却无法正确纠错。然而，当传送出现三个错，使（000）经传送变成了（111），或（111）经传送变成了（000），此时译码器会判定传送无错而造成漏检。这说明，传送出现三个错的情况已经超出了（3，1）重复码的检纠错能力。

同理，读者可自行考察（4，1）重复码和（5，1）重复码。结论是：（4，1）重复码的最小距离 $d=4$，它能纠正单个错误，也可用于检测 3 个错误；（5，1）重复码的最小距离 $d=5$，它能纠正两个错误，或用于检测四个错误。

定理（8.2）是纠错编码理论中最重要的基本定理之一，它说明了码的最小距离 d 与纠错能力之间的关系。由于 d 是码的一个重要参数，因此经常用（n，k，d）表示最小距离为 d 的码。

那么，如何来构造一个距离为 d 的线性分组码呢？

由于线性码满足 $\boldsymbol{cH}^T=\boldsymbol{0}$，码矢 $\boldsymbol{c}=[c_{n-1},c_{n-2},\cdots,c_1,c_0]$为 $1\times n$ 行矢，校验矩阵 $\boldsymbol{H}$是$(n-k)\times n$矩阵，即 $\boldsymbol{H}$由 n 个列矢量组成，记为 $\boldsymbol{H}=\begin{bmatrix}\boldsymbol{h}_{n-1} & \boldsymbol{h}_{n-2} & \cdots & \boldsymbol{h}_0\end{bmatrix}$，则条件 $\boldsymbol{cH}^T=\boldsymbol{0}$ 就等效于

$$c_{n-1}\boldsymbol{h}_{n-1}+c_{n-2}\boldsymbol{h}_{n-2}+\cdots+c_0\boldsymbol{h}_0=\sum_{i=0}^{n-1}c_i\boldsymbol{h}_i=\boldsymbol{0} \tag{8-15}$$

这就是说 $\boldsymbol{H}$中的 n 个列矢是线性相关的，换句话说，$\boldsymbol{H}$中的 n 个列矢其线性组合为零，而以组合系数构成的矢量就是码矢!

若 $\boldsymbol{H}$中任意 l 列线性无关，则在 $\boldsymbol{H}$中总可以找到 l+1 列线性相关。由于这里讨论的是二元码，c_i的取值非 0 即 1，则在码矢 $\boldsymbol{c}=[c_{n-1},\ c_{n-2},\ \cdots,\ c_1,\ c_0]$中，起码应有 l+1 个 c_i=1，才能使$\sum_{i=0}^{n-1}c_i\boldsymbol{h}_i=0$（即 l+1 列相加等于零）。这就意味着码 $\boldsymbol{C}$的最小距离 d 由式（8-16）决定。

$$d=l+1 \tag{8-16}$$

定理 8.3 （n，k）线性分组码的最小距离为 d 的充要条件，是 $\boldsymbol{H}$矩阵中任意 $d-1$ 列线性无关。

定理 8.3 是构造最小距离为 d 的线性分组码的基础。

推论 8.1 （n，k）线性分组码的最小距离 d 的最大可能取值等于 $n-k+1$。

由于（n，k）线性分组码的 $\boldsymbol{H}$矩阵是一个（$n-k$）$\times n$ 阶矩阵，最多只可能有 $n-k$ 列线性无关。若码的最小距离等于 $n-k+1$，则称其为极大最小距离可分码，简称 MDS 码。

【例 8.5】 给定生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&1&0\\0&1&0&0&1&0&1\\0&0&1&0&0&1&1\\0&0&0&1&1&1&1\end{bmatrix}$，生成（7，4）线性分组码，$\boldsymbol{G}$是一个标准生成矩阵，由此可求得一致校验矩阵 $\boldsymbol{H}=\begin{bmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{bmatrix}$，$l$=2，$d$=3，$t=\left\lfloor\dfrac{d-1}{2}\right\rfloor=\dfrac{3-1}{2}=1$，$e\leq d-1=3-1=2$。

给出信息组 $\boldsymbol{u}=0010$，

对应码字 $\boldsymbol{c}=\boldsymbol{uG}=\begin{bmatrix}0&0&1&0\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0&1&1&0\\0&1&0&0&1&0&1\\0&0&1&0&0&1&1\\0&0&0&1&1&1&1\end{bmatrix}=\begin{bmatrix}0&0&1&0&0&1&1\end{bmatrix}$

（1）将 $\boldsymbol{c}$遣入信道发送，假设第 3 位出错，信道输出 $\boldsymbol{y}=0000011$，根据式（8-4）计算伴随式 $\boldsymbol{s}=\boldsymbol{Hy}^T=\begin{bmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{bmatrix}\begin{bmatrix}0\\0\\0\\0\\0\\1\\1\end{bmatrix}=\begin{bmatrix}0\\1\\1\end{bmatrix}$，$\boldsymbol{s}$与 $\boldsymbol{H}$矩阵的第 3 列相同，可判断 $\boldsymbol{y}$的第 3 列

出错，错误图案 $\boldsymbol{e}=0010000$，译码 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}=0000011+0010000=0010011$，估值 $\hat{\boldsymbol{y}}$ 与发送原码 $\boldsymbol{c}$ 相同，可见 1 位错误是可纠的；

（2）假设第 1，2 位出错，信道输出 $\boldsymbol{y}$ = 1110011，计算伴随式 $\boldsymbol{s}=\boldsymbol{H}\boldsymbol{y}^T=\begin{bmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\\0\\1\\1\end{bmatrix}=\begin{bmatrix}0\\1\\1\end{bmatrix}$，此 $\boldsymbol{s}$ 也与 $\boldsymbol{H}$ 矩阵的第 1 列相同。如果译码器仍按原来的规则译码，即对 $\boldsymbol{y}$ 中第 3 位码元实施纠错，结果反而会引起更多的错误，造成错误译码。但 $\boldsymbol{s}\neq\boldsymbol{0}$，说明 $\boldsymbol{y}$ 出错，只是无法判断错在哪几位，可见 2 位错误是可检测的，但无法纠正；

（3）假设第 1，2，3 位同时出错，信道输出 $\boldsymbol{y}$ = 1100011，计算伴随式 $\boldsymbol{s}=\boldsymbol{H}\boldsymbol{y}^T=\begin{bmatrix}1&1&0&1&1&0&0\\1&0&1&1&0&1&0\\0&1&1&1&0&0&1\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\\0\\1\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$，从 $\boldsymbol{s}=\boldsymbol{0}$ 来看 $\boldsymbol{y}$ 没错，可见 3 位（及其以上）错误是不可检测的。

8.4 标准阵列和译码

8.4.1 标准阵列

线性分组码的 n 位码符号由 k 位信息位加上 $n-k$ 位校验位组成，这 n 位码符号取自符号集 $\{a_1,a_2,\cdots,a_q\}$，在整个 n 维空间 $\boldsymbol{V}_n$ 共有 q^n 个矢量。

线性分组码对应 k 维子空间 $\boldsymbol{V}_n^k$，在 k 维子空间中，共有 q^k 个矢量，这 q^k 个矢量就是选用码矢，其余 q^n-q^k 个矢量称为禁用矢量。

下面将 q^n 个矢量排成标准阵列。

第一步：从全零码矢开始，把所有选用码矢 $\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_{q^k}\}$ 依次写成一行；

第二步：选一个在第一行中没列出的禁用矢量 $\boldsymbol{y}_1$ 写在第二行的第一列，其余各列都用 $\boldsymbol{y}_1$ 和第一行对应码矢相加（模 2 加）；

第三步：再选一个第一行、第二行没列出的禁用矢量 $\boldsymbol{y}_2$，其余各列都用 $\boldsymbol{y}_2$ 和第一行对应码矢相加（模 2 加）；

第四步：如此重复，直至把所有 q^n 个矢量列完。

把这样列出的表格称为标准阵列。取 $q=2$，则 2^n 个矢量列出的标准阵列如表 8-7 所示。

表 8-7　线性分组码的标准阵列

许 用 码 字	$\boldsymbol{c}_1$（陪集首）	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	…	$\boldsymbol{c}_{2^k}$
禁用码字	$\boldsymbol{y}_1$	$\boldsymbol{y}_1+\boldsymbol{c}_2$	$\boldsymbol{y}_1+\boldsymbol{c}_3$	…	$\boldsymbol{y}_1+\boldsymbol{c}_{2^k}$
	$\boldsymbol{y}_2$	$\boldsymbol{y}_2+\boldsymbol{c}_2$	$\boldsymbol{y}_2+\boldsymbol{c}_3$	…	$\boldsymbol{y}_2+\boldsymbol{c}_{2^k}$
	⋮	⋮	⋮	⋮	⋮
	$\boldsymbol{y}_{2^{n-k}-1}$	$\boldsymbol{y}_{2^{n-k}-1}+\boldsymbol{c}_2$	$\boldsymbol{y}_{2^{n-k}-1}+\boldsymbol{c}_3$	…	$\boldsymbol{y}_{2^{n-k}-1}+\boldsymbol{c}_{2^k}$

【例 8.6】　二元（5，3）码，生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&1&1\\0&1&0&1&0\\0&0&1&1&1\end{bmatrix}$，信源有 8 个消息待发，对应信源编码器的 8 个输出序列，即

{000，001，010，011，100，101，110，111}

根据 $\boldsymbol{c}=\boldsymbol{u}\boldsymbol{G}$ 编码，得到 8 个码矢，即

{00000，00111，01010，01101，10011，10100，11001，11110}

按上述方式将 2^5=32 个 5 重矢量排成标准阵列，如表 8-8 所示。

第一行为 8 个码矢，其余各行为本行的第一个元素与第一行元素的模 2 加。

表 8-8　将 32 个矢量排成标准阵列

00000	00111	01010	01101	10011	10100	11001	11110
00001	00110	01011	01100	10010	10101	11000	11111
00010	00101	01000	01111	10001	10110	11011	11100
00100	00011	01110	01001	10111	10000	11101	11010

码字 $\boldsymbol{c}=(c_4，c_3，c_2，c_1，c_0)$，通过有噪信道传输，信道输出 $\boldsymbol{y}=(y_4，y_3，y_2，y_1，y_0)$，由于信道噪声干扰，$\boldsymbol{y}$ 序列中的某些码元可能与 $\boldsymbol{c}$ 序列中对应码元的值不同，即传输产生了错误。在二进制序列中，错误无非是 1 错成 0 或 0 错成 1，因此，信道中的干扰可以用二进制序列

$$\boldsymbol{e}=(e_4，e_3，e_2，e_1，e_0)$$

表示，相应于有错的各位 e_i 取值为 1，无错的各位 e_i 取值为 0，则有

$$\boldsymbol{y}=\boldsymbol{c}+\boldsymbol{e} \tag{8-17}$$

称 $\boldsymbol{e}$ 为信道的错误图样。

显然，当 $\boldsymbol{e}=0$ 时，$\boldsymbol{y}=\boldsymbol{c}$，表示接收序列 $\boldsymbol{y}$ 无错；当 $\boldsymbol{e}\neq 0$ 时，$\boldsymbol{y}\neq\boldsymbol{c}$，表示接收序列 $\boldsymbol{y}$ 有错。当 $\boldsymbol{c}$ 序列长为 n 时，信道可能产生的错误图样 $\boldsymbol{e}$ 的数目共有 2^n 种。在此例中 $n=5$，信道可能产生的错误图样 $\boldsymbol{e}$ 的数目共有 2^5=32 种，即表 8-8 中列出的 32 个 5 重矢量都是可能的错误图样，从另一个角度来说，表 8-8 中列出的 32 个 5 重矢量也都是可能的信道输出矢量 $\boldsymbol{y}$。

译码器的任务就是要从收到的 $\boldsymbol{y}$ 中得到 $\hat{\boldsymbol{c}}$，或者说由 $\boldsymbol{y}$ 中解出错误图样 $\boldsymbol{e}$，然后得到 $\hat{\boldsymbol{c}}=\boldsymbol{y}+\boldsymbol{e}$。这里 $\hat{\boldsymbol{c}}$ 是对码字 $\boldsymbol{c}$ 的估值，若 $\hat{\boldsymbol{c}}=\boldsymbol{c}$，则译码正确，否则译码错误。

8.4.2　陪集分解

由标准阵列的构成法，得到结论：

（1）第一行是 q^k 个选用码矢，以后每行都是第一行的陪集，每行的第一个元素称为陪集

首，分别记为 $\boldsymbol{e}_0,\boldsymbol{e}_1,\cdots,\boldsymbol{e}_{q^{n-k}-1}$；

（2）陪集首是前面先列出的元素中没有出现过的，从而该陪集中的元素也是前面没有出现过的；

（3）如果选的陪集首是该行中的另一个元素，则该行中的元素还是原来的 q^k 个元素，只不过排列顺序变了，这说明该行中的每个元素都可以做陪集首；

（4）根据（1）和（2），最后把所有 q^n 个元素全列完了（q^{n-k} 行×q^k 列 $=q^n$ 个元素）；

（5）同一陪集中所有元素的伴随式相同，不同陪集的伴随式不同。

上述（1），（2），（3），（4）条是不言而喻的，下面证明第（5）条。

证明：

① 取第 i 个陪集中的第 j 个元素 $\boldsymbol{e}_i+\boldsymbol{c}_j$，根据式(8-4)，它的伴随式 $\boldsymbol{s}=\boldsymbol{H}(\boldsymbol{e}_i+\boldsymbol{c}_j)^T=\boldsymbol{H}\boldsymbol{e}_i^T$，可见，$\boldsymbol{s}$ 仅与第 i 个陪集首有关，而与 j 的取值无关，故同一陪集的伴随式相同。

② 下面证明不同陪集的伴随式不同，用反证法。

设第 i 个陪集与第 k 个陪集的伴随式相同，即 $\boldsymbol{H}\boldsymbol{e}_i^T=\boldsymbol{H}\boldsymbol{e}_k^T$，则

$$\boldsymbol{H}\left(\boldsymbol{e}_i+\boldsymbol{e}_k\right)^T=\boldsymbol{0} \tag{8-18}$$

式（8-18）说明 $\boldsymbol{e}_i+\boldsymbol{e}_k$ 是一个码矢，即 $\boldsymbol{e}_i+\boldsymbol{e}_k=\boldsymbol{c}\in\boldsymbol{C}$，则

$$\boldsymbol{e}_i=\boldsymbol{e}_k+\boldsymbol{c} \tag{8-19}$$

式（8-19）表明第 k 个陪集中有一个元素 $\boldsymbol{e}_k+\boldsymbol{c}$ 与第 i 个陪集首 $\boldsymbol{e}_i$ 相同，根据标准阵列的排法，这是不可能的，故不同陪集的伴随式不同。

【例 8.7】 在例 8-6 所列的标准阵列中，每行的第一个元素为该行的陪集首，有

$$\boldsymbol{e}_0=00000,\quad \boldsymbol{e}_1=00001,\quad \boldsymbol{e}_2=00010,\quad \boldsymbol{e}_3=00100$$

由矩阵 $\boldsymbol{G}$ 可得 $\boldsymbol{H}=\begin{bmatrix}1&1&1&1&0\\1&0&1&0&1\end{bmatrix}$，根据式（8-4）$\boldsymbol{s}=\boldsymbol{H}\boldsymbol{e}_i^T$，计算出 4 个伴随式，即

$$\boldsymbol{s}_0=00,\quad \boldsymbol{s}_1=01,\quad \boldsymbol{s}_2=10,\quad \boldsymbol{s}_3=11$$

从 $\boldsymbol{s}=\boldsymbol{H}\boldsymbol{y}^T=\boldsymbol{H}(\boldsymbol{e}_i+\boldsymbol{c}_j)^T=\boldsymbol{H}\boldsymbol{e}_i^T$ 来看，若 $\boldsymbol{e}=\boldsymbol{0}$，则 $\boldsymbol{s}=\boldsymbol{0}$；若 $\boldsymbol{e}\neq\boldsymbol{0}$，则 $\boldsymbol{s}\neq\boldsymbol{0}$，伴随式 $\boldsymbol{s}$ 只由错误图样 $\boldsymbol{e}$ 决定，即伴随式 $\boldsymbol{s}$ 是否为全零矢量可以作为判断一个码字传送是否出错的依据。当 $\boldsymbol{s}\neq\boldsymbol{0}$ 时，译码器要做的就是如何从伴随式 $\boldsymbol{s}$ 中找到错误图样 $\boldsymbol{e}$，从而译出发送的码字 $\hat{\boldsymbol{c}}=\boldsymbol{y}+\boldsymbol{e}$。

设（n,k）码的校验矩阵

$$\boldsymbol{H}=\begin{bmatrix}h_{0,n-1}&h_{0,n-2}&\cdots&h_{0,0}\\h_{1,n-1}&h_{1,n-2}&\cdots&h_{1,0}\\\vdots&\vdots&&\vdots\\h_{n-k-1,n-1}&h_{n-k-1,n-2}&\cdots&h_{n-k-1,0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{h}_{n-1}&\boldsymbol{h}_{n-2}&\cdots&\boldsymbol{h}_0\end{bmatrix}$$

式中，$\boldsymbol{h}_{n-j}$ 是 $\boldsymbol{H}$ 矩阵的第 j 列，它是一个 $n-k$ 重列矢量。设码字传送发生 t 位错，且分别为第 j_1，j_2，…，j_t 位有错，则

$$\boldsymbol{e}=\begin{pmatrix}0&\cdots&e_{j_1}&\cdots&0&\cdots&e_{j_2}&\cdots&0&\cdots&e_{j_t}&\cdots&0&\cdots&0\end{pmatrix}$$

在二进制情况下，$e_{j_1}, e_{j_2},\cdots, e_{j_t}$ 为 1，那么伴随式

$$\boldsymbol{s}=\boldsymbol{e}\cdot\boldsymbol{H}^T$$

$$=\begin{bmatrix}0 & \cdots & e_{j_1} & \cdots & 0 & \cdots & e_{j_2} & \cdots & 0 & \cdots & e_{j_t} & \cdots & 0 & \cdots & 0\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{h}_{n-1}\\ \boldsymbol{h}_{n-2}\\ \vdots\\ \boldsymbol{h}_0\end{bmatrix}$$

$$=e_{j_1}\cdot\boldsymbol{h}_{n-j_1}+e_{j_2}\cdot\boldsymbol{h}_{n-j_2}+\cdots+e_{j_t}\cdot\boldsymbol{h}_{n-j_t}$$

$$=\boldsymbol{h}_{n-j_1}+\boldsymbol{h}_{n-j_2}+\cdots+\boldsymbol{h}_{n-j_t} \qquad (8\text{-}20)$$

式（8-20）说明，$\boldsymbol{s}$ 是 $\boldsymbol{H}$ 矩阵中相应于 $e_{j_s}\neq 0$ 的那些列 $\boldsymbol{h}_{n-j_s}$ 的线性组合，显然 $\boldsymbol{s}$ 也是一个 $n–k$ 重矢量。

【例 8.8】 以表 8-4 给出的（7, 4）线性分组码为例。已知该码的校验矩阵 $\boldsymbol{H}$ 为

$$\boldsymbol{H}=\begin{bmatrix}1 & 1 & 1 & 0 & 1 & 0 & 0\\ 1 & 1 & 0 & 1 & 0 & 1 & 0\\ 1 & 0 & 1 & 1 & 0 & 0 & 1\end{bmatrix}$$

（1）若传送时发生一位错误

$\boldsymbol{e}_1$=（1000000），根据式（8-4）计算伴随式得 $\boldsymbol{s}_1=\boldsymbol{H}\boldsymbol{e}_1^T=\begin{bmatrix}1\\1\\1\end{bmatrix}$

$\boldsymbol{s}$ 正好就是 $\boldsymbol{H}$ 矩阵的第 1 列，在一般情况下，若传送时发生 1 位错码，如果错的是第 j 列，则伴随式 $\boldsymbol{s}$ 恰好就等于矩阵的第 j 列。

（2）若传送时发生二位错误

设 $\boldsymbol{e}_2$=（0001001），第 4 位和第 7 位出错，计算伴随式得

$$\boldsymbol{s}_2=\boldsymbol{H}\boldsymbol{e}_2^T=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

又设 $\boldsymbol{e}_3$=（1000001），第 1 位和第 7 位出错，计算伴随式得 $\boldsymbol{s}_3=\boldsymbol{H}\boldsymbol{e}_3^T=\begin{bmatrix}0\\0\\1\end{bmatrix}$

观察这两种情况，伴随式 $\boldsymbol{s}_2$ 和 $\boldsymbol{s}_3$ 均不为 0，说明传送的码字有错。然而，错误图样 $\boldsymbol{e}_2$ 和 $\boldsymbol{e}_3$ 虽然不同，但它们对应的伴随式 $\boldsymbol{s}_2$ 和 $\boldsymbol{s}_3$ 却完全相同，因此无法判定到底是哪两位出错，从而不能纠正错误。经观察还可以发现，$\boldsymbol{s}_2$ 是 $\boldsymbol{H}$ 矩阵中第 4 列与第 7 列之和，$\boldsymbol{s}_3$ 是 $\boldsymbol{H}$ 矩阵中第 1 列与第 7 列之和。

（3）若传送时发生三位错误

设 $\boldsymbol{e}_4$ =（0111000），第 2 位、第 3 位和第 4 位出错，计算伴随式得 $\boldsymbol{s}_4=\boldsymbol{H}\boldsymbol{e}_4^T=\begin{bmatrix}0\\0\\0\end{bmatrix}$

这种情况下伴随式 $\boldsymbol{s}_4$=**0**，表明无错，实际情况是发生了误码，说明误码三位（及其以上）检测不出来，注意观察可以发现，$\boldsymbol{s}_4$ 是 $\boldsymbol{H}$ 矩阵中第 2 列、第 3 列和第 4 列之和。一般来说，伴随式 $\boldsymbol{s}$ 是 $\boldsymbol{H}$ 矩阵中相应于码字出错位置所对应的列矢量的线性组合。

事实上，通过观察 $\boldsymbol{H}$ 矩阵也可算得 l=2，d=3，$t=\left\lfloor \dfrac{d-1}{2} \right\rfloor = \dfrac{3-1}{2}=1$，$e \leq d-1=3-1=2$。

8.4.3 译码

1．用标准阵列译码

接收到 $\boldsymbol{y}$ 后，到标准阵列中去找（因为 q^n 个矢量全部列在其中，总可以找到），假设在某列中找到了 $\boldsymbol{y}$，则把它相应地译成该列的第一个矢量。

【例 8.9】 仍考虑例 8.6 所述的二元（5，3）码，假设接收矢量为 $\boldsymbol{y}$=10110，在表 8-8 列出的标准阵列中，找到 $\boldsymbol{y}$ 位于第 6 列，相应地译成 10100。错误图案是 $\boldsymbol{y}$=10110 所在行的陪集首 $\boldsymbol{e}_2=00010$。

利用标准阵列译码，需要把 2^n 个 n 重矢量存在译码器中，译码器的复杂性将随 n 成指数增加，当 n 较大时难以使用。

2．用伴随式译码

在例 8.6 中讲过，在标准阵列中出现的 q^n 个矢量都是可能的接收矢量，它们具有 $\boldsymbol{y}=\boldsymbol{e}_j+\boldsymbol{c}_j$ 的形式，将 $\boldsymbol{y}$ 看成是发送码矢 $\boldsymbol{c}_j$ 时收到的矢量，则 $\boldsymbol{e}_j$ 就是错误图案，根据最小错误译码准则，可译码如下。

在标准阵列的每一行中选重量最轻的矢量作为陪集首（例 8.6 二元（5，3）码的标准阵列就是如此），接收到 $\boldsymbol{y}$ 后，计算伴随式 $\boldsymbol{s}=\boldsymbol{H}\boldsymbol{y}^{\mathrm{T}}$，再在 $\boldsymbol{s}$ 所对应的陪集中，找到陪集首 $\boldsymbol{e}$，译码输出 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}$。

第一行为码矢，从第二行开始，每一列与本列的第一行元素都只相差一个陪集首，在例 8-6 中讲过，在标准阵列中出现的 q^n 个矢量都是可能的错误图案，从 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}$ 可看出，只有陪集首 $\boldsymbol{e}$ 才是可纠正的错误图案，而陪集首选的是本行中重量最轻的矢量，所以 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}$ 就是最小距离译码，即最小错误概率译码。

综上所述，我们可以将标准阵列简化成更为实用的译码表。译码表保留了标准阵列中的 2^{n-k} 个可纠正错误图样 $\boldsymbol{e}_j$（即陪集首）与其伴随式 $\boldsymbol{s}=\boldsymbol{H}\cdot\boldsymbol{e}_j^{\,T}$ 之间的一一对应关系，译码器存储该表后，在译码时就可以查表实现从伴随式到错误图样的转换。表 8-9 所示就是表 8-8 所列（5，3）线性分组码标准阵列的译码表。

表 8-9 （5，3）线性分组码标准阵列的译码表

伴随式 $\boldsymbol{s}$	错误图样（陪集首）$\boldsymbol{e}$
000	00000
001	00001
010	00010
011	00100

【例 8.10】 仍考虑例 8.6 所述的二元（5，3）码，假设接收矢量为 $\boldsymbol{y}$=10110，

计算出伴随式 $\boldsymbol{s}=\boldsymbol{H}\boldsymbol{y}^T=\begin{bmatrix}1&1&1&1&0\\1&0&1&0&1\end{bmatrix}\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$，找到对应陪集首 $\boldsymbol{e}_2$=00010，

译码输出 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}=10110\oplus00010=10100$，$\hat{\boldsymbol{y}}$ 就是 $\boldsymbol{y}$ 所在列的第一个矢量。

这种译码方法与利用标准阵列译码的结果是等效的。

用上述方法译码时，译码正确的概率与陪集首的选择有关。根据最大后验概率译码准则，重量最轻的错误图样产生的可能性最大，所以应该优先选择重量小的 n 重作为陪集首。这样构造的译码表，使得 $\boldsymbol{e}_j+\boldsymbol{c}_i$ 与 $\boldsymbol{c}_i$ 之间的距离最小，从而使译码器能以更大的正确概率译码，这就是最小距离译码。在 BSC 信道中，它等效于最大后验概率译码。

将表 8-8 所示的译码表存入译码器，只需要存储 2^{n-k} 个 n 重（陪集首）及 2^{n-k} 个 $n-k$ 重（伴随式）矢量。近年来，由于大规模集成电路的发展，存储器的容量大大增加，而价格日益下降；又由于正确选择了陪集首，使译码表译码的错误概率较小，计算伴随式 $\boldsymbol{s}$ 及由 $\boldsymbol{s}$ 找 $\boldsymbol{e}$ 均较为容易，译码步骤较为简单，因此这种译码方法有相当好的应用前景。

3．二元线性码的误码率

所有 q^n 个矢量都是可能出现的错误图案，根据伴随式译码方法，选重量最轻的矢量作为陪集首，采用最小距离译码准则时，陪集首项的集合就是可纠正图案的集合，即错成 $\boldsymbol{y}=\boldsymbol{c}+\boldsymbol{e}$ 时，可纠正为 $\hat{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{e}=\boldsymbol{c}+\boldsymbol{e}+\boldsymbol{e}=\boldsymbol{c}$，则正确译码的概率就是陪集首出现的概率。

设 $\omega(l)$ 是第 l 个陪集首的重量，则对于二进制对称信道 BSC，有

$$p_c=\sum_{l=1}^{2^{n-k}}p^{w(l)}(1-p)^{n-w(l)}=\sum_{i=0}^{n}\alpha_i p^i(1-p)^{n-i}$$

式中，α_i 是重量为 i 的陪集首个数，则错误概率为

$$p_e=1-p_c=1-\sum_{i=0}^{n}\alpha_i p^i(1-p)^{n-i} \tag{8-21}$$

【例 8.11】 在例 8.6 的二元（5，3）码中，$\alpha_0=1,\ \alpha_1=3,\ \alpha_2=\alpha_3=\alpha_4=0$，则

$$p_c=(1-p)^5+3p(1-p)^4$$

即

$$p_e=1-p_c=1-(1-p)^5+3p(1-p)^4$$

8.5 汉 明 码

汉明码是 1950 年由汉明（R.W.Hamming）提出的能纠正单个错误的线性分组码。它性能良好，既具有较高的可靠性，又具有较高的传输效率，而且编译码电路较为简单，易于工程实现，因此汉明码在发现后不久，就得到了广泛的应用。

8.5.1 汉明码的构造

我们的目的是要寻找一个能纠正单个错误，且信息传输率（即码率 $r=k/n$）最大的线性分组码。我们已经知道，具有纠正单个错误能力的线性分组码的最小距离应为 3，即要求其 $\boldsymbol{H}$ 矩

阵中至少任意两列线性无关。要做到这一点，只要 $\boldsymbol{H}$ 矩阵满足“两无”——无相同的列，无全零列就可以了。

(n, k) 线性分组码的 $\boldsymbol{H}$ 矩阵是一个 $(n-k)\times n = r\times n$ 阶矩阵，这里 $r = n-k$ 是校验元的数目。显然，r 个校验元能组成 2^r 列互不相同的 r 重矢量，其中非全零矢量有 2^r-1 个。如果用这 2^r-1 个非全零矢量作为 $\boldsymbol{H}$ 矩阵的全部列，即令 $\boldsymbol{H}$ 矩阵的列数 $n=2^r-1$，则此 $\boldsymbol{H}$ 矩阵的各列均不相同，且无全零列，由此可构造一个纠正单个错误的 (n, k) 线性分组码。

同时，2^r-1 是 n 所能取的最大值，因为如果 $n>2^r-1$，那么 $\boldsymbol{H}$ 矩阵的 n 列中必会出现相同的两列，这样就不能满足对 $\boldsymbol{H}$ 矩阵的要求。而由于 $n = 2^r-1$ 是 n 所能取的最大值，也就意味着码率 r 取得了最大值，即

$$r = \frac{k}{n} = \frac{n-r}{n} = 1-\frac{r}{n} = 1-\frac{r}{2^r-1} \tag{8-22}$$

这样设计出来的码是符合我们的要求的，这样的码就是汉明码。

定义 8.6 若 $\boldsymbol{H}$ 矩阵的列由非全零且互不相同的所有二进制 r 重矢量组成，则由此得到的线性分组码称为 GF(2)上的 $(2^r-1, 2^r-1-r)$ 汉明码。

表 8-10 列出了几种 r 取不同值时汉明码的 $(n=2^r-1, k=2^r-1-r)$ 值。

表 8-10 几种汉明码的 (n, k) 值

r	$n=2^r-1$	$k=2^r-1-r$
1	1	0
2	3	1
3	7	4
4	15	11
5	31	26
6	127	121

【例 8.12】 取 $r=3$，构造 GF(2)上的（7，4）汉明码。

当 $r=3$ 时，有 7 个非全零的三重矢量：

（001），（010），（011），（100），（101），（110），（111）

构成 $\boldsymbol{H}$ 矩阵

$$\boldsymbol{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

由此得到一个能纠正单错的（7，4）汉明码。若码字传输中左边第一位出错，则相应的伴随式 $\boldsymbol{s}$=（001）就是 $\boldsymbol{H}$ 矩阵的第一列，也正好是“1”的二进制表示。同理可知，无论哪一位出错，它对应的伴随式就是该位的二进制表示，故译码十分方便，特别适用于计算机内部运算和记忆系统中的纠错。

如要得到系统码形式的 $\boldsymbol{H}$ 矩阵，只需对上述矩阵进行初等变换交换列即可。

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

相应地，生成矩阵 $\boldsymbol{G}$ 为

$$\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&1&1\\0&1&0&0&1&1&0\\0&0&1&0&1&0&1\\0&0&0&1&0&1&1\end{bmatrix}$$

由此构成的（7，4）汉明码已示于表 8-4。

8.5.2 汉明限与完备码

根据 8.3.4 节译码部分的介绍可知，任何一个二进制（n，k）线性分组码，若要纠正 t 个错误，则应使小于或等于 t 个错误所组成的所有错误图样，都必须有不同的伴随式与之对应，即以下不等式成立：

$$2^{n-k}\geq\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{t}=\sum_{i=0}^{t}\binom{n}{i} \tag{8-23}$$

式中，2^{n-k} 为全部 $r=n-k$ 重矢量数目，即伴随式数目，$\sum_{i=0}^{t}\binom{n}{i}$ 为所有错误个数小于或等于 t 的错误图样数。

式（8-23）称为汉明限。该限是构造任何二进制码所必须满足的，也就是构造码的必要条件。

如果某一（n, k）线性分组码能使式（8-23）中等号成立，即错误图样总数正好等于伴随式数目，则称这种码为完备码。完备码相当于在标准阵列中，能将重量小于等于 t 的所有错误图样作为陪集首，而大于 t 的错误图样都不作为陪集首，其校验元得到了充分的利用。显然无论 r 取何值，汉明码正是可纠 t=1 位错误的完备码。

如果一个（n, k）线性分组码，除了能将重量小于等于 t 的所有错误图样作为陪集首外，还有部分（但不是全部）重量大于 t 的错误图样作为陪集首，则称这种码为准完备码。

表 8-8 列出的二元（5，3）码，就不是一个完备码，由它的生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&1&1\\0&1&0&1&0\\0&0&1&1&1\end{bmatrix}$，可得到校验矩阵 $\boldsymbol{H}=\begin{bmatrix}1&1&1&1&0\\1&0&1&0&1\end{bmatrix}$，由此算出 l=1，d=2，$t=\left\lfloor\frac{d-1}{2}\right\rfloor=\left\lfloor\frac{2-1}{2}\right\rfloor=0$，纠错能力为 $t=0$，除全零码矢外，标准阵列中还有部分重量等于 1 的矢量作为陪集首。

【例 8.13】 （7，4）系统码，生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&1&1\\0&1&0&0&1&1&0\\0&0&1&0&1&0&1\\0&0&0&1&0&1&1\end{bmatrix}$，校验矩阵 $\boldsymbol{H}=\begin{bmatrix}1&1&1&0&1&0&0\\1&1&0&1&0&1&0\\1&0&1&1&0&0&1\end{bmatrix}$，由此算出 l=2，3=2，$t=\left\lfloor\frac{d-1}{2}\right\rfloor=\left\lfloor\frac{3-1}{2}\right\rfloor=1$，列出它的标准阵列如表 8-11 所示。

表 8-11 （7，4）系统码的标准阵列

0000000	0001011	0010101	0011110	0100110	0101101	0110011	0111000	1000111	1001100	1010010	1011001	1100001	1101010	1110100	1111111
0000001	0001010	0010100	0011111	0100111	0101100	0110010	0111001	1000110	1001101	1010011	1011000	1100000	1101011	1110101	1111110
0000010	0001001	0010111	0011100	0100100	0101111	0110001	0111010	1000101	1001110	1010000	1011011	1100011	1101000	1110110	1111101
0000100	0001111	0010001	0011010	0100010	0101001	0110111	0111100	1000011	1001000	1010110	1011101	1100101	1101110	1110000	1111011
0001000	0000011	0011101	0010110	0101110	0100101	0111011	0110000	1001111	1000100	1011010	1010001	1101001	1100010	1111100	1110111
0010000	0011011	0000101	0001110	0110110	0111101	0100011	0101000	1010111	1011100	1000010	1001001	1110001	1111010	1100100	1101111
0100000	0101011	0110101	0111110	0000110	0001101	0100011	0011000	1100111	1101100	1110010	1111001	1000001	1001010	1010100	1011111
1000000	1001011	1010101	1011110	1100110	1101101	1110011	1111000	0000111	0001100	0010010	0011001	0100001	0101010	0110100	0111111

从表 8-11 可看出，上述二元（7，4）码是完备汉明码。

本 章 小 结

线性分组码是最简单、最基本的一类纠错码，它是学习纠错码的基础。本章的主要内容是：

（1）纠错码的基本概念：信道纠错编码及其目的、差错控制系统的三种实现方式、检错码与纠错码、分组码与卷积码、随机错误与突发错误。

（2）线性分组码的编码：线性分组码的定义、生成矩阵和校验矩阵的构成、系统码的生成矩阵和校验矩阵、对偶码、编码实现电路。

（3）线性分组码的检纠错能力：汉明距离和汉明重量、码的最小距离、码的最小距离与检纠错能力的关系、距离为 d 的线性分组码的构造。

（4）伴随式与译码：错误图样、接收序列的伴随式、码的标准阵列的构成、选择陪集首、将标准阵列简化为译码表、线性分组码的译码步骤。

（5）汉明码：汉明码的定义、构造一个汉明码、汉明限、完备码与准完备码。

本章的重点是线性分组码的构成理论及其编译码方法，应结合实例搞清概念。

思考题与习题

8.1 什么是检错码？什么是纠错码？两者有什么不同？

8.2 试述分组码的概念，并说明分组码的码率 r 的意义。

8.3 什么是码的生成矩阵和校验矩阵？一个（n，k）线性分组码的生成矩阵和校验矩阵各是几行几列的矩阵？

8.4 什么样的码称为系统码？系统码的生成矩阵和校验矩阵在形式上有何特点？

8.5 什么是对偶码？试举例说明之。

8.6 试述码的距离和重量的概念。线性分组码的最小距离有何实际意义？

8.7 如果要构造一个能纠两个错的线性分组码，则其 $\boldsymbol{H}$ 矩阵中至少应保证多少列线性无关？

8.8 什么是接收序列 $\boldsymbol{y}$ 的伴随式 $\boldsymbol{s}$？为什么伴随式 $\boldsymbol{s}$ 只由错误图样 $\boldsymbol{e}$ 决定？

8.9 如何构造一个码的标准阵列？标准阵列有什么性质？

8.10　如何利用标准阵列译码？为什么说用标准阵列译码时，译码错误概率的大小与陪集首的选择有关？

8.11　计算例 8.6 所给出的二元（5，3）码的纠错能力 t 和检错能力 e，结果与“陪集首项的集合就是可纠正图案的集合”这一说法是否矛盾？

8.12　什么是完备码？为什么说汉明码是完备码？

8.13　某分组码的校验矩阵

$$\boldsymbol{H}=\begin{bmatrix}0&1&1&1&0&0\\1&0&1&0&1&0\\1&1&0&0&0&1\end{bmatrix}$$

求：（1）n=? k=? 该码的码字有多少？

（2）该码的生成矩阵；

（3）矢量 010111 和 100011 是否码字？

8.14　某二元（n，k）系统线性分组码的全部码字如下：

00000　　01011　　10110　　11101

求：（1）n=? k=?

（2）码的生成矩阵 $\boldsymbol{G}$和校验矩阵 $\boldsymbol{H}$。

8.15　已知一个线性分组码的校验矩阵

$$\boldsymbol{H}=\begin{bmatrix}0&1&1&1&1&0&0\\1&0&1&1&0&1&0\\1&1&0&1&0&0&1\end{bmatrix}$$

试求其生成矩阵。当输入信息序列为 1001 1100 1101 时，求编码器输出的码字序列。

8.16　构造题 8.15 中（7，4）分组码的对偶码，构造其系统码形式的 $\boldsymbol{G}$矩阵和 $\boldsymbol{H}$矩阵，并写出全部码字。

8.17　某（5，2）线性分组码的 $\boldsymbol{H}$矩阵为

$$\boldsymbol{H}=\begin{bmatrix}1&1&1&0&0\\1&0&0&1&0\\1&1&0&0&1\end{bmatrix}$$

求：（1）该码的 $\boldsymbol{G}$矩阵；

（2）码的标准阵列；

（3）码的简化译码表；

（4）该码是否完备码？

8.18　试构造 GF(2)上的（15，11）汉明码，求出其系统码形式的 $\boldsymbol{H}$矩阵和 $\boldsymbol{G}$矩阵。

8.19　证明最小距离为 7 的二进制（23，12）Golay 码是完备码。

8.20　设一个（7，4）分组码的生成矩阵为

$$\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&1&1\\0&1&0&0&1&0&1\\0&0&1&0&0&1&1\\0&0&0&1&1&1&0\end{bmatrix}$$

求：（1）该码的全部码字；

（2）码的标准阵列；

（3）码的简化译码表。

8.21　信源的 4 个消息$\{a_1, a_2, a_3, a_4\}$被编成 4 个码长为 5 的二元码字 00000，01101，10111，11010 发送。

（1）试给出码的生成矩阵 G 和校验矩阵 H；

（2）若上述码字通过固有误码率为 $p<0.5$ 的二进制对称信道传输，请列出其标准矩阵，并计算错误概率 p_e。

8.22　设二元（6，3）的生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&1\\0&1&0&0&0&1\\0&0&1&1&1&0\end{bmatrix}$，给出它的校验矩阵 $\boldsymbol{H}$，并计算该码的最小距离。

8.23　设α是域 GF(2^5)的本原元，令 $\hat{\alpha}=\alpha^3$，若码的生成多项式以 $\hat{\alpha}$，$\hat{\alpha}^2$，$\hat{\alpha}^3$，$\hat{\alpha}^4$ 为根，求出该码的生成多项式 $g(x)$、校验多项式 $h(x)$、码长 n 和信息位数 k。

8.24　证明任何（n，k）线性码的码字重量全部为偶数或偶数重量和奇数重量各占一半。

8.25　信息序列 $\boldsymbol{u}=u_0u_1$，将其编为三元（4，2）码 $\boldsymbol{c}=c_0c_1c_2c_3$，编码规则如下：

$$c_0=u_0，\ c_1=u_1，\ c_2=u_0\oplus u_1，\ c_3=u_0\oplus 2u_1$$

（1）写出此码的生成矩阵 G 和校验矩阵 H；

（2）求出此码的纠错能力 t 和检错能力 e；

（3）写出所有可纠正图案的伴随式。

8.26　一个（8，4）系统码的编码规则如下：

$c_0=u_0$，$c_1=u_1$，$c_2=u_2$，$c_3=u_3$，$c_4=u_1\oplus u_2\oplus u_3$，$c_5=u_0\oplus u_1\oplus u_2$，$c_6=u_0\oplus u_1\oplus u_3$，$c_7=u_0\oplus u_2\oplus u_3$

（1）写出此码的生成矩阵 G 和校验矩阵 H；

（2）计算该码的最小距离 d。

8.27　二元（7，4）的标准阵列如表 8-11 所示，如果接收矢量为 $\boldsymbol{y}=0011100$，试根据表 8-11，利用伴随式和标准阵列两种方法译码，并比较译码结果。

第 9 章
循　环　码

内容提要

循环码是线性分组码中一个重要的子类。本章首先提出了循环码的定义以及循环码的多项式描述方法，给出了生成多项式和校验多项式的定义，论述了循环码构成的有关重要定理；接着讨论了循环码的编译码方法及其实现电路，最后介绍了已获得广泛应用的循环汉明码、BCH 码等。

知识要点

循环码，生成多项式，校验多项式、（n–k）级编码器，k 级编码器，Meggit 译码器，BCH 码。

教学建议

循环码在纠错码理论中具有重要的地位，自 1957 年普兰奇（N.Prange）提出循环码后，其结构、性质及编译码方法得到了迅速和详尽的研究，取得了许多重要成果。循环码之所以受到如此重视，是因为它具有完整的代数结构，这种码的代数结构完全建立在有限域基础上，具有很多有用的性质，由于其代数结构和线性反馈移位寄存器的数学结构相同，故它们的编译码可以方便地利用移位寄存器实现。可以说，循环码是目前研究得最成熟的一类码。现在大多数有实用价值的纠错码都属于循环码的范畴。建议学时数为 6 学时，重点学习循环码的构成理论及其编译码方法。为了便于学生理解，讲授时应结合实例。

9.1 循环码的一般概念

9.1.1 循环码的定义

【例 9.1】 （7，4）线性分组码，生成矩阵 $\boldsymbol{G}=\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$，这是一个标准生成矩阵，将其生成的全部 16 个码矢按重量分类列于表 9-1。

表 9-1 将（7，4）码按重量分类

重量=0	重量=3	重量=4	重量=7
0000000	0001011	0011101	1111111
	0010110	0111010	
	0101100	1110100	
	1011000	1101001	
	0110001	1010011	
	1100010	0100111	
	1000101	1001110	

观察表 9-1 所示的所有码字，可以发现码字的循环特性：在表中的 16 个码字中，重量为 3 的 7 个码字形成一个循环，重量为 4 的 7 个码字形成另一个循环，全“0”码字和全“1”码字可以看成自身循环。设想将任一码字中的 7 个码元排在一个圆周上，则从圆周的任一码元开始，按顺时针方向移动一周，都将构成该码的一个码字。这就是循环码的由来。

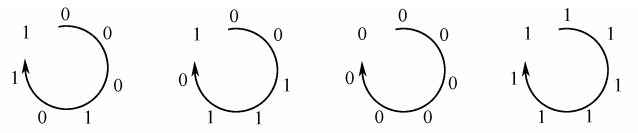


图 9-1 （7，4）线性码的码字循环图

定义 9.1 一个（n，k）线性码 $\boldsymbol{C}$，若对任意 $\boldsymbol{c}=(c_{n-1},c_{n-2},\cdots,c_0)\in\boldsymbol{C}$，将码矢中的各码符号循环左移（或右移）一位，恒有 $\boldsymbol{c}'=(c_{n-2},\ \cdots\ ,c_0,c_{n-1})\in\boldsymbol{C}$，就称 $\boldsymbol{C}$ 为（n，k）循环码。

由于（n, k）线性分组码是 n 维线性空间 $\boldsymbol{V}_n$ 中的一个 k 维子空间，因此（n，k）循环码是 n 维线性空间 $\boldsymbol{V}_n$ 中的一个 k 维循环子空间。

9.1.2 循环码的多项式描述

为了借助代数这一工具研究循环码，可以将一个码矢 $\boldsymbol{c}=(c_{n-1},c_{n-2},\cdots,c_0)$ 中的各码元 c_i（i=0,1,…, n−1）看成一个多项式的系数，从而将码矢 $\boldsymbol{c}$ 表示成码多项式的形式。

码多项式

$$c(x) = c_{n-1}x^{n-1}+c_{n-2}x^{n-2}+\cdots+c_1x+c_0 \tag{9-1}$$

这里，码多项式的系数就是码矢中各码元的值，式中，x^i（i =0,1,…, n−1）的幂次 i 起到了标明码矢 $\boldsymbol{c}$ 中各分量 c_i 所处的位置的作用。例如 x^{n-1} 表示对应的系数 c_{n-1} 是码矢 $\boldsymbol{c}$ 的第 n−1 个分量，x^i 表示对应的系数 c_i 是 $\boldsymbol{c}$ 的第 i 个分量。

这样定义了码多项式后，在带模余数运算下，就可以利用第 7 章学过的关于多项式域的各种知识来处理码多项式。

码的循环移位可用代数表示，即

$$x\cdot c(x)\ [\mathrm{mod}\ (x^n-1)] = (c_{n-1}x^n+c_{n-2}x^{n-1}+\cdots+c_1x^2+c_0x)\ [\mathrm{mod}\ (x^n-1)]$$
$$= c_{n-2}x^n+\cdots+c_1x^2+c_0x+c_{n-1}$$

即循环码的一位循环移位可由模 x^n−1 下的码多项式 $c(x)$乘以 x 的运算给出。

【例 9.2】 给定二元循环码 $\boldsymbol{C}$中的码字 $\boldsymbol{c}$ =11010110，对应码多项式 $c(x) = x^7+x^6+x^4+x^2+x$，计算 $x\cdot c(x)\ [\mathrm{mod}\ (x^7-1)]$。

取 $c(x)$的系数进行竖式运算：

$$\begin{array}{r} 1 \\ 10000001\overline{)\,11010110} \\ \underline{10000001} \\ 1010111 \end{array}$$

即 $x\cdot c(x)\ [\mathrm{mod}\ (x^7-1)]= x^6+x^4+x^2+x+1$，对应码矢 $\boldsymbol{c}'$=01010111，可看出 $\boldsymbol{c}'$ 是 $\boldsymbol{c}$ 的循环左移。

后面给出的多项式，都要对其进行模 x^n−1 运算，不再重复写出。

9.2 循环码的生成多项式和生成矩阵

9.2.1 生成多项式

定理 9.1 设 $g(x)$是码 $\boldsymbol{C}$中的最低次码多项式，则 $\boldsymbol{C}$是循环码的充要条件是：所有非零码多项式都是 $g(x)$的倍式。

证明：设$\partial^\circ g(x) = m$。

① 必要性。设 $\boldsymbol{C}$是循环码，则根据循环码定义，$xg(x)$，$x^2g(x)$，…，$x^{n-m-1}g(x)$都是码多项式，由线性码的封闭性知，它们的线性组合$(a_{n-m-1}x^{n-m-1} + a_{n-m-2}x^{n-m-2}+\cdots+ a_1x + a_0)\ g(x)$ 也是码多项式，即 $g(x)$ 的倍式是码多项式。

将 $\boldsymbol{C}$ 中的任一码多项式 $v(x)$ 写成如下形式：$v(x) = a(x)\ g(x) + r(x)$（设$\partial^\circ\ r(x) < \partial^\circ\ g(x)$），因为 $g(x)$的倍式 $a(x)\ g(x)$是码多项式，则 $v(x) + a(x)\ g(x) = r(x)$ 也是码多项式（封闭性），由于 $\partial^\circ\ r(x) < \partial^\circ\ g(x)$，与 $g(x)$ 是最低阶次码多项式矛盾，故 $r(x)$必为 $\boldsymbol{0}$（全零码矢），说明 $\boldsymbol{C}$中的码都可写成 $g(x)$的倍式。

② 充分性。设 $\boldsymbol{C}$中的码 $r(x)$是 $g(x)$的倍式，记为

$$v(x) = a(x)g(x) = (a_{n-m-1}x^{n-m-1} + a_{n-m-2}x^{n-m-2}+\cdots+ a_1x + a_0)\ g(x) \tag{9-2}$$

对式（9-2）两边乘以 x，再取$[\mathrm{mod}\ (x^n-1)]$，得

$$xv(x) = (a_{n-m-2}x^{n-m-1} + a_{n-m-3}x^{n-m-2}+\cdots+ a_0x + a_{n-m-1})\ g(x) \tag{9-3}$$

式（9-3）右边仍为 $g(x)$的倍式，根据前面的结论，$g(x)$的倍式是码多项式，所以左边也是

码多项式，说明任一码多项式 $v(x)$的循环位移仍是码多项式，由循环码的定义知 $\boldsymbol{C}$是循环码。

证毕

定理 9.1 说明，只要找到 $g(x)$，就可由它生成循环码，称 $g(x)$为循环码的**生成多项式**。

那么 $g(x)$是否存在且唯一呢？

定理 9.2 在（n，k）循环码 $\boldsymbol{C}$中，有唯一的最低次首一码多项式。

证明：（反证法）假设 $\begin{cases} g(x)=x^m+g_{m-1}x^{m-1}+\cdots+g_1x+g_0 \\ g'(x)=x^m+g'_{m-1}x^{m-1}+\cdots+g'_1x+g'_0 \end{cases}$ 都是 $\boldsymbol{C}$的最低次首一码多项式，则由线性码的封闭性知：

$$g(x)+g'(x)=(g_{m-1}+g'_{m-1})x^{m-1}+\cdots+(g_1+g'_1)x+(g_0+g'_0)$$

也是 $\boldsymbol{C}$的一个码多项式，从上式可看出：$\partial^\circ[g(x)+g'(x)]<\partial^\circ g(x)$，这与 $g(x)$是最低次多项式不符，说明假设错误，$g(x)+g'(x)$只能是全零码矢），说明假设错误。

$g(x)$如何找？下面两条定理说明了 $g(x)$的特性。

定理 9.3 （n，k）循环码的生成多项式 $g(x)$，其幂次为 $n-k$，且常数项必不为零。

证明：因为 $c(x)=a(x)g(x)$，$\partial^\circ c(x)=n-1$，设$\partial^\circ g(x)=m$，则$\partial^\circ a(x)=n-m-1$，即 $c(x)$可以写成

$$c(x)=(a_{n-m-1}x^{n-m-1}+a_{n-m-2}x^{n-m-2}+\cdots+a_1x+a_0)g(x) \qquad (9\text{-}4)$$

系数 $a_0, a_1, \cdots, a_{n-m-1}$取自 GF($q$)，则根据式（9-4）可构成 q^{n-m}个码多项式 $c(x)$，又由线性分组码的构成知道，信息位为 k 的码含 q^k个码矢，即有 $q^k=q^{n-m}\Rightarrow m=n-k$ 。

记 $g(x)=x^m+g_{m-1}x^{m-1}+\cdots+g_1x+g_0$，下面证明常数项 g_0必不为零。

反证法：设常数项 $g_0=0$，$g(x)=x^m+g_{m-1}x^{m-1}+\cdots+g_2x^2+g_1x$，将 $g(x)$循环移位 $n-1$次，得

$$\begin{aligned} x^{n-1}g(x)&=(x^{n-1+m}+g_{m-1}x^{n-1+m-1}+\cdots+g_2x^{n-1+2}+g_1x^{n-1+1}) \quad [\mathrm{mod}\,(x^n-1)] \\ &=x^{m-1}+g_{m-1}x^{m-2}+\cdots+g_2x+g_1 \end{aligned}$$

$x^{n-1}g(x)$仍是码多项式，由上式可看出$\partial^\circ[x^{n-1}g(x)]<\partial^\circ g(x)$，这一结果与 $g(x)$是最低次码多项式相矛盾，故假设不成立，$g(x)$的常数项必不为零。

证毕

定理 9.3 说明，（n，k）循环码的生成多项式为

$$g(x)=x^{n-m}+g_{n-m-1}x^{n-m-1}+\cdots+g_1x+g_0 \qquad (9\text{-}5)$$

式（9-5）中 $g_0\neq0$，通常我们研究二元域 GF(2)，则 $g_0=1$。

线性分组码讨论的多项式都是在模 x^n-1 下的余式运算，下面一条定理指出了生成多项式 $g(x)$与 x^n-1 的关系。

定理 9.4 生成多项式 $g(x)$必整除 x^n-1。

证明：

记

$$x^n-1=a(x)g(x)+r(x) \qquad (9\text{-}6)$$

式（9-6）中$\partial^\circ r(x)<\partial^\circ g(x)$，对式（9-6）两边取[mod (x^n-1)]，得

$$0=[a(x)g(x)+r(x)]\ [\mathrm{mod}\,(x^n-1)]$$

即 $r(x)=[a(x)g(x)]\ [\mathrm{mod}\,(x^n-1)]$。因为 $g(x)$的倍式都是码多项式，故 $r(x)$是码多项式；又因为$\partial^\circ r(x)<\partial^\circ g(x)$，与 $g(x)$是最低次码多项式相矛盾，故 $r(x)=0$，说明 $g(x)$ 整除 x^n-1。

证毕

综合定理 9.1 至定理 9.4 可知，构造（n，k）循环码的的问题在于分解因式 x^n-1，找到 $n-k$ 次多项式 $g(x)$，$g(x)$就是（n，k）循环码的生成矩阵，将 GF(q)上的所有 q^k 个（$k-1$）次信息组多项式与 $g(x)$相乘，就得到 q^k 个码多项式 $c(x)$。

【例 9.3】 构造（7，4）循环码：要构造一个（7，4）循环码，就是要在 x^7-1 中找出一个 $n-k=3$ 次的因式 $g(x)$作为码的生成多项式，逐一用所有信息多项式作为乘因子，它们与 $g(x)$的倍式就构成了（7，4）循环码。

分解因式 $x^7-1=(x^3+x+1)(x^3+x^2+1)(x+1)$，前面两个因式都是 $n-k=3$ 次多项式，都可作为生成多项式 $g(x)$，此例中选 $g(x)=(x^3+x+1)$，其所生成的循环码如表 9-2 所示。

表 9-2 由 $g(x)$ 生成码多项式

信 息 位	信息多项式	$g(x)$的倍式	码 多 项 式	码 矢
0000	0	$0\cdot(x^3+x+1)$	0	0000000
0001	1	$1\cdot(x^3+x+1)$	x^3+x+1	0001011
0010	x	$x\cdot(x^3+x+1)$	x^4+x^2+x	0010110
0011	$x+1$	$(x+1)\cdot(x^3+x+1)$	$x^4+x^3+x^2+1$	0011101
0100	x^2	$(x^2)\cdot(x^3+x+1)$	$x^5+x^3+x^2$	0101100
0101	x^2+1	$(x^2+1)\cdot(x^3+x+1)$	x^5+x^2+x+1	0100111
0110	x^2+x	$(x^2+x)\cdot(x^3+x+1)$	$x^5+x^4+x^3+x$	0111010
0111	x^2+x+1	$(x^2+x+1)\cdot(x^3+x+1)$	x^5+x^4+1	0110001
1000	x^3	$x^3\cdot(x^3+x+1)$	$x^6+x^4+x^3$	1011000
1001	x^3+1	$(x^3+1)\cdot(x^3+x+1)$	x^6+x^4+x+1	1010011
1010	x^3+x	$(x^3+x)\cdot(x^3+x+1)$	$x^6+x^3+x^2+x$	1001110
1011	x^3+x+1	$(x^3+x+1)\cdot(x^3+x+1)$	x^6+x^2+1	1000101
1100	x^3+x^2	$(x^3+x^2)\cdot(x^3+x+1)$	$x^6+x^5+x^4+x^2$	1110100
1101	x^3+x^2+1	$(x^3+x^2+1)\cdot(x^3+x+1)$	$x^6+x^5+x^4+x^3+x^2+x+1$	11111111
1110	x^3+x^2+x	$(x^3+x^2+x)\cdot(x^3+x+1)$	x^6+x^5+x	1100010
1111	x^3+x^2+x+1	$(x^3+x^2+x+1)\cdot(x^3+x+1)$	$x^6+x^5+x^3+1$	1101001

从表 9-2 中可看出，除全零码矢（0000000）和全 1 码矢外（1111111），其余码矢可分别由码矢（0001011）和（0011101）循环移位得到。

【例 9.4】 构造（7，3）循环码。分解因式 $x^7-1=(x^3+x+1)(x^3+x^2+1)(x+1)$，分别取 $g_1(x)=(x^3+x+1)(x+1)=x^4+x^3+x^2+1$ 和 $g_2(x)=(x^3+x^2+1)(x+1)=x^4+x^2+x+1$ 作为生成多项式，生成两个不同的（7，3）码列于表 9-3。

表 9-3 分别由生成多项式 $g_1(x)$ 和 $g_2(x)$生成的（7,3）循环码

$g_1(x)=x^4+x^3+x^2+1$		$g_2(x)=x^4+x^2+x+1$	
码字多项式	码 字	码字多项式	码 字
$g(x)=x^4+x^3+x^2+1$	0011101	$g(x)=x^4+x^2+x+1$	0010111
$xg(x)=x^5+x^4+x^3+x$	0111010	$xg(x)=x^5+x^3+x^2+x$	0101110
$x^2g(x)=x^6+x^5+x^4+x^2$	1110100	$x^2g(x)=x^6+x^4+x^3+x^2$	1011100
$(x^2+1)g(x)=x^6+x^5+x^3+1$	1101001	$(x+1)g(x)=x^5+x^4+x^3+1$	0111001

续表

$g_1(x)=x^4+x^3+x^2+1$		$g_2(x)=x^4+x^2+x+1$	
码字多项式	码　字	码字多项式	码　字
$(x^2+x+1)g(x)=x^6+x^4+x+1$	1010011	$(x^2+x)g(x)=x^6+x^5+x^4+x$	1110010
$(x+1)g(x)=x^5+x^2+x+1$	0100111	$(x^2+x+1)g(x)=x^6+x^5+x^2+1$	1100101
$(x^2+x)g(x)=x^6+x^3+x^2+x$	1001110	$(x^2+1)g(x)=x^6+x^3+x+1$	1001011
$0g(x)=0$	0000000	$0g(x)=0$	0000000

从表 9-3 中可看出，除全零码矢（0000000）外，$g_1(x)$生成的码可由码矢（0011101）循环移位得到，$g_2(x)$生成的码可由码矢（0010111）循环移位得到。

9.2.2　生成矩阵

仔细观察表 9-3 可看出，（7，3）循环码的 8 个码字可由 $g(x)$, $x\,g(x)$, $x^2\,g(x)$的线性组合产生，这三个码多项式是线性无关的，它们组成了码的一组基底，生成（7，3）循环码。根据定义 8.2，这组基底所构成的 3×7 阶矩阵就是上述（7，3）循环码的生成矩阵，用 $\boldsymbol{G}$ 表示，以 $g_1(x)= x^4+ x^3+ x^2+1$ 生成的码为例，

$$\boldsymbol{G}_1=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\0&0&1&1&1&0&1\end{bmatrix}$$

对于一般的（n，k）循环码，设其生成多项式为

$$g(x)=g_{n-k}x^{n-k}+g_{n-k-1}x^{n-k-1}+\cdots+g_1x+g_0$$

由于 $g(x), x\,g(x), \cdots, x^{k-1}g(x)$共 k 个码多项式（它们表示分别将 $g(x)$循环移位 0 次，1 次，…，k–1 次）必线性无关，故可用它们组成码的一组基底，而与这些码多项式相对应的 k 个线性无关的码矢就构造出 $k\times n$ 阶生成矩阵 $\boldsymbol{G}$，即

$$\boldsymbol{G}=\begin{bmatrix}g_{n-k}&g_{n-k-1}&\cdots&&g_1&g_0&0&\cdots&&0\\0&g_{n-k}&g_{n-k-1}&\cdots&&g_1&g_0&0&\cdots&0\\\vdots&&&&&&&&&\\0&\cdots&0&g_{n-k}&g_{n-k-1}&&\cdots&&g_1&g_0\end{bmatrix}\qquad(9\text{-}7)$$

【例 9.5】　GF(2)中的（7，4）循环码，$x^7-1=(x^3+x+1)(x^3+x^2+1)(x+1)$，取 $g(x)=(x^3+x+1)$，则生成矩阵为 $\boldsymbol{G}=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$，由 $g(x)$或 $\boldsymbol{G}$都可生成（7，4）循环码。

① 由 $g(x)$生成：根据 $c(x)=u(x)\,g(x)$，其生成码多项式如表 9-2 所示。

② 由 $\boldsymbol{G}$生成：根据 $\boldsymbol{c}=\boldsymbol{u}\,\boldsymbol{G}$，其生成码矢如表 9-2 所示。

例如，信息组 $\boldsymbol{u}=1001$，编码为

$$\boldsymbol{c}=\boldsymbol{u}\boldsymbol{G}=\begin{bmatrix}1 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 1 & 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 1 & 0 & 1 & 1\end{bmatrix}=\begin{bmatrix}1 & 0 & 1 & 0 & 0 & 1 & 1\end{bmatrix}$$

其余码矢也可类似计算。

9.3 循环码的校验多项式和校验矩阵

1．反多项式

多项式 $a(x)=a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+\cdots+a_1x+a_0$ 的反多项式定义为

$$\begin{aligned}a^*(x)&=x^{n-1}a(x)\\&=x^{n-1}[a_{n-1}x^{-(n-1)}+a_{n-2}x^{-(n-2)}+\cdots+a_1x^{-1}+a_0]\\&=a_{n-1}+a_{n-2}x+\cdots+a_1x^{n-2}+a_0x^{n-1}\\&=a_0x^{n-1}+a_1x^{n-2}+\cdots+a_{n-2}x+a_{n-1}\end{aligned}$$

定理 9.5 $a(x)b(x)=0\ [\mathrm{mod}\ (x^n-1)]$的充要条件是 $a(x)$的系数矢量与 $x^i b^*(x)(i=1,2,\cdots,n)$的系数矢量正交。

证明：

① 必要性。记 $a(x)=a_{n-1}x^{n-1}+\cdots+a_1x+a_0$，$b(x)=b_{n-1}x^{n-1}+\cdots+b_1x+b_0$

则 $a(x)\,b(x)=(a_{n-1}x^{n-1}+\cdots+a_1x+a_0)(b_{n-1}x^{n-1}+\cdots+b_1x+b_0)$

$$\begin{aligned}&=(a_{n-1}b_0+\cdots+a_1b_{n-2}+a_0b_{n-1})x^{n-1}+(a_{n-1}b_{n-1}+\cdots+a_1b_{n-3}+a_0b_{n-2})x^{n-2}+\cdots+\\&\quad(a_{n-1}b_1+\cdots+a_1b_{n-1}+a_0b_0)x^0\end{aligned}$$

$a(x)b(x)=0$，就要求上式中 x 各幂次的系数等于零，这就意味着 $b^*(x)$不论如何移位，与 $a(x)$的系数总是正交的。

② 充分性。将上述证明过程反推回去，可证得充分性。

证毕

2．校验多项式、校验矩阵

已知 $g(x)$整除 x^n-1，记为

$$x^n-1=g(x)\,h(x) \tag{9-8}$$

$$\partial^\circ g(x)=n-k,\quad \partial^\circ h(x)=k$$

对式（9-8）两边取 $\mathrm{mod}\ (x^n-1)$，则有 $0=[g(x)\,h(x)]\ [\mathrm{mod}\ (x^n-1)]$。

由定理 9.5 知，$g(x)$ 与 $x^i h^*(x)$（$i=0,1,\cdots,n-k-1$）正交，另一方面由式（8-7）有 $\boldsymbol{G}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$，即生成矩阵与校验矩阵应该是正交的，故取多项式 $x^i h^*(x)$（$i=0,1,\cdots,n-k-1$）的系数为行矢可构成校验矩阵

$$\boldsymbol{H}=\begin{bmatrix}h_k^* & h_{k-1}^* & \cdots & h_0^* & 0 & \cdots & 0\\ 0 & h_k^* & h_{k-1}^* & \cdots & h_0^* & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & 0 & h_k^* & h_{k-1}^* & \cdots & h_0^*\end{bmatrix}$$

由于 $\boldsymbol{H}$是由 $h(x)$的反多项式 $h^*(x)$得到的，相应地称 $h(x)$为校验多项式。

【例 9.6】　GF(2)中的（7，4）循环码，$x^7-1=(x^3+x+1)(x^3+x^2+1)(x+1)$

（1）若取生成多项式为　　$g_1(x)=x^3+x+1$

则校验多项式　　$h_1(x)=(x^7-1)/g_1(x)=x^4+x^2+x+1$

$h_1(x)$的反多项式　　$h_1^*(x)=x^4(x^{-4}+x^{-2}+x^{-1}+1)=x^4+x^3+x^2+1$

校验矩阵　　$\boldsymbol{H}_1=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\0&0&1&1&1&0&1\end{bmatrix}$

可验算 $\boldsymbol{G}_1\boldsymbol{H}_1^T=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\0&1&1\\1&0&1\\0&1&0\\0&0&1\end{bmatrix}=\boldsymbol{0}$

（2）若取生成多项式为　$g_2(x)=x^3+x^2+1$

则校验多项式　　$h_2(x)=(x^7-1)/g_2(x)=x^4+x^3+x^2+1$

$h_2(x)$的反多项式　　$h_2^*(x)=x^4(x^{-4}+x^{-3}+x^{-2}+1)=x^4+x^2+x+1$

校验矩阵　　$\boldsymbol{H}_2=\begin{bmatrix}1&0&1&1&1&0&0\\0&1&0&1&1&1&0\\0&0&1&0&1&1&1\end{bmatrix}$

由 $g_1(x)$和 $g_2(x)$生成的码多项式列于表 9-4。

表 9-4　分别由生成码多项式 $g_1(x)$ 和 $g_2(x)$生成的（7,4）循环码

信　息　位	信息多项式	由 $g_1(x)$生成码多项式	由 $g_2(x)$生成码多项式
0000	0	0	0
0001	1	x^3+x+1	x^3+x^2+1
0010	x	x^4+x^2+x	x^4+x^3+x
0011	$x+1$	$x^4+x^3+x^2+1$	x^4+x^2+x+1
0100	x^2	$x^5+x^3+x^2$	$x^5+x^4+x^2$
0101	x^2+1	x^5+x^2+x+1	x^5+x^2+x+1
0110	x^2+x	$x^5+x^4+x^3+x$	$x^5+x^4+x^3+1$
0111	x^2+x+1	x^5+x^4+1	x^5+x+1
1000	x^3	$x^6+x^4+x^3$	$x^6+x^5+x^3$
1001	x^3+1	x^6+x^4+x+1	$x^6+x^5+x^2+1$
1010	x^3+x	$x^6+x^3+x^2+x$	$x^6+x^5+x^2+x$
1011	x^3+x+1	x^6+x^2+1	$x^6+x^5+x^4+x^3+x^2+x+1$
1100	x^3+x^2	$x^6+x^5+x^4+x^2$	$x^6+x^4+x^3+x^2$
1101	x^3+x^2+1	$x^6+x^5+x^4+x^3+x^2+x+1$	x^6+x^4+1
1110	x^3+x^2+x	x^6+x^5+x	x^6+x^2+x
1111	x^3+x^2+x+1	$x^6+x^5+x^3+1$	x^6+x^3+x+1

由表 9-4 可看出，由不同的生成矩阵所生成的码是不同的。

由于循环码也归属于线性码，故同样可利用初等变换将生成矩阵 $\boldsymbol{G}$ 变换成标准形式的生成矩阵，这样在（n，k）循环码中，由标准形式的生成矩阵所生成的码字，前面 k 位是信息位，后面 $n-k$ 位是校验位。

【例 9.7】 在例 9.6 中，可通过初等变换将 $\boldsymbol{G}_1$ 变换成标准形式，即

$$\boldsymbol{G}_1=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}\xrightarrow{\text{初等变换}}\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

相应地校验矩阵也变换为标准形式，即 $\boldsymbol{H}_1=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{bmatrix}$

其所生成的循环码如表 9-5 所示。

表 9-5 由 uG 生成码矢

信 息 位	码矢 $\boldsymbol{uG}$	信 息 位	码矢=$\boldsymbol{uG}$
0000	0000000	1000	1000101
0001	0001011	1001	1001110
0010	0010110	1010	1010011
0011	0011101	1011	1011000
0100	0100111	1100	1100010
0101	0101100	1101	1101001
0110	0110001	1110	1110100
0111	0111010	1111	1111111

从表中可看出，除全零码矢（0000000）和全 1 码矢（1111111）外，其余码矢可分别由码矢（0001011）和（0011101）循环移位得到。

综上所述，不难得到如下结论：

① （n，k）循环码 $\boldsymbol{C}$ 的生成多项式 $g(x)$是码 $\boldsymbol{C}$ 中阶次最低的唯一的首一多项式，其阶次 $\partial^{\circ}g(x)=n-k$ 恰好是码字中校验元的数目；

② 生成多项式 $g(x)$是多项式 x^n-1 的因式。要构造一个（n，k）循环码，就要在 x^n-1 的因式中找一个阶次为 $n-k$ 的首一多项式 $g(x)$，逐一用所有信息多项式作为乘因子，它们与 $g(x)$ 的倍式就构成了（n，k）循环码，反之，循环码的每个码字多项式必是以信息多项式作为乘因子的 $g(x)$的倍式；

③ 由 $x^n-1=g(x)h(x)$，$h(x)$称为校验多项式。对于任意一个（n，k）循环码，必有

$$g(x)h(x)=0 \qquad [\bmod (x^n-1)]$$

④ $g(x)$，$xg(x)$，…，$x^{k-1}g(x)$这 k 个码多项式组成循环码 $\boldsymbol{C}$ 的一组基底，而与这 k 个码多项式相对应的 k 个线性无关的码矢构造出 $k\times n$ 阶生成矩阵 $\boldsymbol{G}$；记 $h(x)$的反多项式为 $h^*(x)$，则 $h^*(x)$，$x\,h^*(x)$，…，$x^{n-k-1}\,h^*(x)$这 $n-k$ 个线性无关的矢量构造出$(n-k)\times n$ 阶校验矩阵 $\boldsymbol{H}$，且满足

$$\boldsymbol{G}\cdot\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$$

最后值得指出的是，循环码是线性分组码的一个子类。因此，所有线性分组码的性质均适用于循环码。

9.4 循环码的编码

9.4.1 利用$g(x)$实现编码

如上所述，当循环码的生成多项式 $g(x)$确定后，码就完全确定了。现在讨论当生成多项式 $g(x)$给定以后，如何实现系统循环码的编码问题。

系统码可写成如下形式

$$c = \underbrace{u_{k-1}u_{k-2}\cdots u_1u_0}_{k\text{位信息位}} \quad \underbrace{r_{n-k-1}r_{n-k-2}\cdots r_1r_0}_{(n-k)\text{位校验位}}$$

所以系统循环码的任一码矢可写成码多项式

$$c(x) = x^{n-k}u(x) + r(x) \tag{9-9}$$

$\partial^\circ r(x) = n-k-1$，对式（9-9）两边取[mod $g(x)$]，因为任一码多项式 $c(x)$都是生成多项式 $g(x)$的倍式，所以取模后等式左边为零，即有

$$r(x) = x^{n-k}u(x) \ [\text{mod } g(x)] \tag{9-10}$$

式（9-10）表明，将信息多项式 $u(x)$ 移位 $n-k$ 次，然后用 $g(x)$去除它，余数就是检验多项式 $r(x)$，由此可得编码步骤：

（1）用 x^{n-k} 乘 $u(x)$；

（2）用 $g(x)$除 $x^{n-k}u(x)$，得余式 $r(x)$；

（3）组合 $x^{n-k}u(x) + r(x)$，得码字 $c(x)$。

除法电路可用反馈移位寄存器构成。

1．除法电路

讨论二元域上多项式除法运算的实现电路。

设 GF(2)上的两个多项式

$$a(x) = a_kx^k+a_{k-1}x^{k-1}+\cdots+a_1x+a_0$$

$$b(x) = b_mx^m+b_{m-1}x^{m-1}+\cdots+b_1x+b_0$$

式中，$a(x)$是被除式，$b(x)$是除式。$a_i,b_j\in$GF(2)，$i=0,1,\cdots,k$；$j=0,1,\cdots,m$。

可用图 9-2 所示的电路完成 $a(x)$除以 $b(x)$的运算。图中电路的主体是移位寄存器 D_{m-1}，D_m，…，D_2，D_1，其数目为 m 个，等于除式 $b(x)$的最高幂次，⊕ 表示异或逻辑运算。

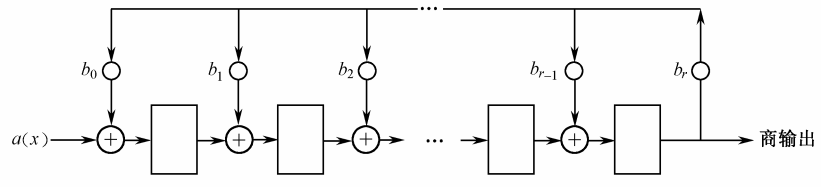


图 9-2　除法电路的一般形式

【例 9.8】　设被除式 $a(x)=x^4+x+1$，除式 $b(x)=x^3+x^2+1$，完成两个多项式相除的运算。

多项式的系数运算：

$$
\begin{array}{r}
\quad\;\; 1\;\;1 \\
1\;1\;0\;1\;\overline{)\,1\;0\;0\;1\;1} \\
\quad\;\; 1\;1\;0\;1 \\
\hline
\quad\;\; 1\;0\;0\;1 \\
\quad\;\; 1\;1\;0\;1 \\
\hline
\quad\;\; 1\;0\;0
\end{array}
$$

得　　$a(x) = (x+1)b(x)+x^2$

实现以上除法运算的除法电路如图 9-3 所示，图中 D_2，D_1，D_0 代表移位寄存器，⊕ 表示异或逻辑运算。

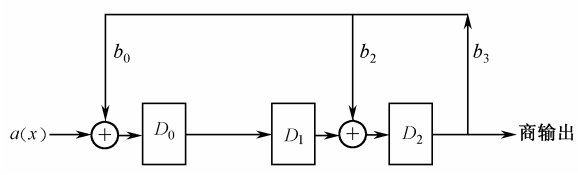


图 9-3　以 $b(x) = x^3+x^2+1$ 为除式的除法电路

表 9-6 给出了除法电路的运算过程。除法电路的运算是按时钟脉冲的节拍进行的。运算开始前所有的触发器清零。随着第一个时钟节拍的到来，运算开始，被除式 $a(x)$的系数（$a_4\ a_3\ a_2\ a_1\ a_0$）=（1 0 0 1 1）按节拍依次输入，先输入 a_4，最后输入 a_0，移位寄存器内容也随输入而不断改变。运算结束时，输出的商为（0 0 0 1 1），即 $x+1$，移位寄存器中的内容（$D_2\ D_1\ D_0$）=（1 0 0），即余式为 x^2。

表 9-6　除法电路运算过程

时 钟 节 拍	输入 $a(x)$	移位寄存器内容			商　输　出
		D_0	D_1	D_2	
0	0	0	0	0	0
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	0
4	1	0	0	1	1
5	1	0	0	1	1

2．编码电路

下面通过一个实例给出循环码利用 $g(x)$实现编码的电路。

【例 9.9】　GF(2)中的（7，4）循环码，生成矩阵 $g(x) = x^3+x+1$，编码电路如图 9-4 所示。反馈抽头按 $g(x)$的系数接，$\begin{cases} g_i = 1 & \text{接通} \\ g_i = 0 & \text{断开} \end{cases} i = 0, 1, \cdots, n-k$，这种结构的电路需要 $n-k$ 个移位寄存器 D。

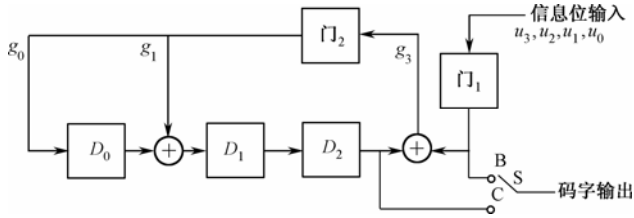


图 9-4　（7，4）循环码编码电路（按 $g(x)$编码）

电路工作过程：

（1）寄存器 $D_0D_1\,D_2$ 预清零；

（2）在前 $k=4$ 个节拍时，门 1、门 2 开，开关 S 接至 B 端，信息位 u_3, u_2, u_1, u_0 由高至低逐位移入寄存器，同时输出到信道，待 4 个节拍结束，寄存器中就是 3 位校验位；

（3）在后 $n-k=3$ 个节拍时，门 1、门 2 关，开关 S 接至 C 端，将寄存器中的校验位依次输出到信道。

例如：对于信息组（1101），对应信息多项式 $u(x)=x^3+x^2+1$，余式

$$r(x)=x^3\ (x^3+x^2+1)\ [\bmod g(x)\ ]=1$$

则根据式（9-9），码多项式

$$c(x)=x^3\,u(x)+r(x)=x^6+x^5+x^3+1$$

对应码矢 $\boldsymbol{c}=\underbrace{1101}_{\text{4 位信息位}}\ \ \underbrace{001}_{\text{3 位校验位}}$。

将校验位的计算过程列于表 9-7。

表 9-7　校验位计算过程

节　拍	输　入	D_0	D_1	D_2
0		0	0	0
1	1	1	1	0
2	1	1	0	1
3	0	1	0	0
4	1	1	0	0

9.4.2　利用$h(x)$实现编码

根据定理 9.1，循环码 $\boldsymbol{C}$ 中的所有非零码多项式 $c(x)$都是生成多项式 $g(x)$的倍式，记为

$$c(x)=m(x)g(x) \tag{9-11}$$

$\partial^\circ g(x)=n-k,\ \partial^\circ m(x)=k-1$，用校验多项式 $h(x)$乘式（9-11）两边，得

$$\begin{aligned}c(x)h(x)&=g(x)\,h(x)\,m(x)\\&=(x^n-1)\,m(x)\\&=x^n m(x)-m(x)\end{aligned} \tag{9-12}$$

由于$\partial^\circ\ x^n m(x)=n+k-1$，对式（9-12）两边取$[\bmod (x^n-1)]$，可看出等式右边的阶次为 $k-1$，即 x^k，…，x^{n-1} 的系数为零，那么等式左边 x^k，…，x^{n-1} 的系数也为零。

$$\begin{aligned}c(x)\,h(x)&=(\,c_{n-1}x^{n-1}+\ \cdots\ +c_2x^2+c_1x+c_0)\,(\,h_k x^k+\ \cdots\ +h_1x+h_0\,)\\&=c_0\,h_0+(c_0\,h_1+c_1\,h_0)\,x+(c_0\,h_2+c_1\,h_1+c_2\,h_0)\,x^2+\ \cdots\end{aligned} \tag{9-13}$$

式（9-13）中 x^l 的系数为 $\sum_{i=0}^{l} h_i c_{l-i}$，$l=0,1,2,\cdots,n-1$，但 x^n，…，x^{n-1+k} 的系数不能如此计算，例如，x^{n-1+k} 的系数为 $c_{n-1}h_k$，由于这些系数与我们讨论的编码无关，故不予考虑。

因为 x^k，…，x^{n-1} 的系数为零，故

$$k \leq l \leq n-1 \text{ 时，} \sum_{i=0}^{l} h_i c_{l-i} = 0 \tag{9-14}$$

令 $l=n-j$，则式（9-14）变为

$$1 \leq j \leq n-k \text{ 时，} \sum_{i=0}^{n-j} h_i c_{n-i-j} = 0 \tag{9-15}$$

由于$\partial^\circ h(x)=k$，$h(x)=h_k x^k+h_{k-1}x^{k-1}+\cdots+h_1 x+h_0$，在式（9-15）的和式中变量 i 的取值最大为 k，另外由于我们讨论的是二元码，有 h_k=1，由式（9-15）得

$$h_k c_{n-k-j} + \sum_{i=0}^{k-1} h_i c_{n-i-j} = 0 \qquad 1 \leq j \leq n-k$$

$$c_{n-k-j} = \sum_{i=0}^{k-1} h_i c_{n-i-j} \qquad 1 \leq j \leq n-k \tag{9-16}$$

由式（9-16）可看出，给定 k 位信息数字 $c_{n-1}, c_{n-2},\cdots, c_{n-k}$，根据式（9-16）就可决定码 $\boldsymbol{c}$ 中的 $n-k$ 位校验数字 $c_{n-k-1}, c_{n-k-2},\cdots, c_0$，如下所示：

$$\begin{cases} j=1 & c_{n-k-1} = \sum_{i=0}^{k-1} h_i c_{n-1-i} = h_0 c_{n-1} + h_1 c_{n-2} + \cdots + h_{k-1} c_{n-k} \\ j=2 & c_{n-k-2} = \sum_{i=0}^{k-1} h_i c_{n-2-i} = h_0 c_{n-2} + h_1 c_{n-3} + \cdots + h_{k-1} c_{n-k-1} \\ & \vdots \\ j=n-k & c_0 = \sum_{i=0}^{k-1} h_i c_{n-(n-k)-i} = \sum_{i=0}^{k-1} h_i c_{k-i} = h_0 c_k + h_1 c_{k-1} + \cdots + h_{k-1} c_1 \end{cases}$$

【例 9.10】 GF(2)中的（7，4）循环码，生成多项式 $g(x)=x^3+x+1$，校验多项式 $h(x)=\dfrac{x^7-1}{x^3+x+1}=x^4+x^2+x+1$，按 $h(x)$编码，编码电路如图 9-5 所示。

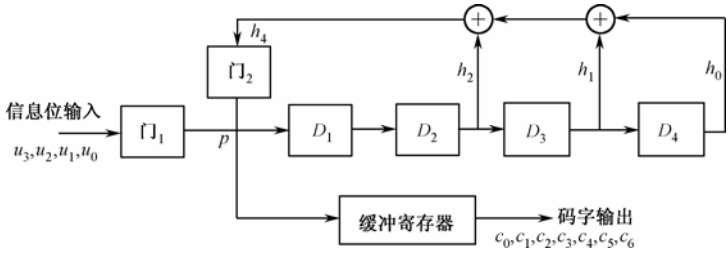


图 9-5 （7，4）循环码编码电路（按 $h(x)$编码）

反馈抽头按 $h(x)$的系数接，$\begin{cases} h_i=1 & \text{接通} \\ h_i=0 & \text{断开} \end{cases}$ $i=0,1,\cdots,k$，这种结构的电路需要 k 个移位寄存器 D。

电路工作过程：

（1）移位寄存器 D_1，D_2，D_3，D_4 及缓冲寄存器清零；

（2）门 1 开，门 2 开，$k=4$ 位信息数字 $c_6=u_3$, $c_5=u_2$, $c_4=u_1$, $c_3=u_0$，逐位移入寄存器 $D_1D_2D_3D_4$，同时输出到缓冲寄存器；

（3）门 1 关，门 2 开，在 p 点得到第一个校验数字 $c_2=c_6h_0+c_5h_1+c_4h_2+c_3h_3$；

（4）移位寄存器移位一次，c_2 移入 D_1，且输出到缓冲寄存器，于是在 p 点又得到第二个校验数字 $c_1=c_5h_0+c_4h_1+c_3h_2+c_2h_3$；

（5）反复进行步骤（3），每移位一次输出一位校验数字至缓冲寄存器，同时在 p 点产生一位新的校验数字，直到 3 个校验数字全部移入缓冲寄存器；

（6）门 2 关，门 1 开，准备接收新的数字。

例如，对于信息位（1101），即 $c_6, c_5, c_4, c_3=1101$，可算出

$$\begin{cases} c_2 = h_0c_6 + h_1c_5 + h_2c_4 + h_3c_3 = 0 \\ c_1 = h_0c_5 + h_1c_4 + h_2c_3 + h_3c_2 = 0 \\ c_0 = h_0c_4 + h_1c_3 + h_2c_2 + h_3c_1 = 1 \end{cases}$$

码矢 $\boldsymbol{c}=\underbrace{1101}_{4\text{位信息位}}\ \underbrace{001}_{3\text{位校验位}}$，与例 9.9 中按 $g(x)$计算的结果一致，将编码过程列于表 9-8。

表 9-8　例 9.9 编码过程

节　　拍	输　　入	$D_1D_2D_3D_4$	缓冲寄存器
门 1 开，门 2 关 0		0000	0000000
1	1	1000	1000000
2	1	1100	1100000
3	0	0110	0110000
4	1	1011	1011000
门 1 关，门 2 开 5	无输入	0101	0101100
6		0010	0010110
7		1001	1001011

例 9.9 利用生成多项式 $g(x)$实现编码的电路采用 $n-k$ 个寄存器，故称为 $n-k$ 级编码器。例 9.10 利用校验多项式 $h(x)$实现编码的电路采用 k 个寄存器，故称为 k 级编码器。一般来说，如果信息元位数 k 大于校验元位数 $r=n-k$，所以用 $n-k$ 级编码器为宜；反之，则以用 k 级编码器为宜。

9.5　循环码的译码

9.5.1　伴随式计算

当码字 $\boldsymbol{c}$ 通过噪声信道传送时，会受到干扰而产生错误。如信道产生的错误图样是 $\boldsymbol{e}$，译码器收到的接收矢量是 $\boldsymbol{y}$，则有

$$\boldsymbol{y}=\boldsymbol{c}+\boldsymbol{e}$$

上式也可写成多项式形式，即

$$y(x) = c(x)+e(x) \tag{9-17}$$

用生成多项式 $g(x)$除以接收多项式 $y(x)$，得

$$y(x) = a(x)g(x)+s(x) \tag{9-18}$$

式中，$a(x)$为商式，$s(x)$为余式。

由于码多项式 $c(x)$是生成多项式 $g(x)$的倍式，即 $c(x) = m(x)g(x)$，比较式（9-17）和式（9-18）可看出，$s(x)$由错误多项式 $e(x)$所决定，与码多项式 $c(x)$无关，因此称 $s(x)$为校验式或伴随多项式，当码字在信道传输中没有发生误码时 $s(x) = 0$。

循环码具有循环移位特性，这使得其伴随式 $s(x)$也有了循环特性，以下就是伴随式 $s(x)$计算电路的一个重要性质。

定理 9.6　设 $s(x)$是接收码字多项式 $y(x)$的伴随式，则 $y(x)$的一次循环移位 $xy(x)[\bmod(x^n-1)]$的伴随式 $s^{(1)}(x)$，是 $s(x)$在伴随式计算电路中无输入时右移一位的结果（称为自发运算），即

$$s^{(1)}(x) = xs(x) \qquad [\bmod g(x)] \tag{9-19}$$

证明：由伴随式定义有

$$s(x) = y(x) \qquad [\bmod g(x)] \tag{9-20}$$

且　　$xs(x) = xy(x) \qquad [\bmod g(x)]$

又由伴随式定义　　$s^{(1)}(x) = xy(x) \qquad [\bmod g(x)]$

比较以上两式得　　$s^{(1)}(x) = xs(x) \qquad [\bmod g(x)]$　　证毕

上述定理可以推广到更一般的情况：对于任何 $i = 1$，2，…，$n-1$，$y(x)$的 i 次循环移位 $x^iy(x)[\bmod g(x)]$的伴随式 $s^{(i)}(x)$，必有

$$s^{(i)}(x) = x^is(x) \qquad [\bmod g(x)]$$

即 $s^{(i)}(x)$是 $s(x)$在伴随式计算电路中无输入时，右移 i 位的结果。

上述性质在循环码的译码中非常有用！

根据式（9-20），伴随式 $s(x)$是接收多项式 $y(x)$关于生成多项式 $g(x)$求模的结果，可以用 9.4.1 节中介绍的除法电路来计算 $s(x)$。

【例 9.11】　取 $g(x) = x^4+x^3+x^2+1$ 为生成多项式，生成（7，3）循环码（见表 9-3），可算出该码能纠正一位错误。

用图 9-6 所示的除法电路来计算伴随式。

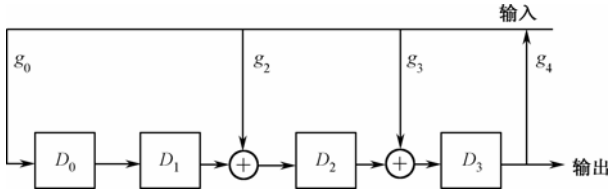


图 9-6　（7，3）循环码的伴随式计算电路

设传送出现一位错，错误图样 $\boldsymbol{e}$=（0 0 0 0 1 0 0），即 $e(x) = x^2$。

由式（9-20）$s(x) = y(x)[\bmod g(x)]$，而 $y(x) = c(x)+e(x)$，又由于码多项式 $c(x)$是生成多项式 $g(x)$的倍式，因此有

$$s(x) = e(x)[\bmod g(x)] \tag{9-21}$$

在此例中，$s(x) = e(x)[\bmod g(x)] = x^2\ [\bmod (x^4+x^3+x^2+1)] = x^2$，即 $\boldsymbol{s}$=（0 1 0 0），将电路计算 $\boldsymbol{s}$ 的过程列于表 9-9（节拍 0～7）。

表 9-9　计算 $\boldsymbol{s}$=0100

节　拍	输　入	D_0	D_1	D_2	D_3
0		0	0	0	
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	1	1	0	1	1
6	0	0	1	0	1
7	0	0	0	1	0
8	无输入右移一位	0	0	0	1

又设错误图样 $\boldsymbol{e}_1$=（0 0 0 1 0 0 0），即 $e^{(1)}(x) = xe(x) = x^3$，相应的伴随式 $s^{(1)}(x) = x^3$ [mod $(x^4+x^3+x^2+1)$] $= x^3+x^2+1$，即 $\boldsymbol{s}_1$=（1 0 0 0），将电路计算 $\boldsymbol{s}$ 的过程列于表 9-10。

表 9-10　计算 s_1 = 1000

节　拍	输　入	D_0	D_1	D_2	D_3
0		0	0	0	
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	1	1	0	1	1
5	0	0	1	0	1
6	0	0	0	1	0
7	0	0	0	0	1

可以验证，$\boldsymbol{s}_1$ 是 $\boldsymbol{s}$ 在图 9.5 所示的除法电路中无输入时，右移一位的结果，也即自发运算的结果，见表 9-9（节拍 8）。

9.5.2　循环码的纠错译码

译码器的任务就是从 $y(x)$中得到估值错误图样 $\hat{e}(x)$，然后求得估值码字 $\hat{c}(x) = y(x) + \hat{e}(x)$，并从中得到信息组 $\hat{m}(x)$。

循环码的译码可按以下三个步骤进行：

（1）由接收到的 $y(x)$计算伴随式多项式 $s(x)$；

（2）根据伴随式 $s(x)$找出对应的估值错误图样 $\hat{e}(x)$；

（3）计算 $\hat{c}(x) = y(x) + \hat{e}(x)$，得到估值码字 $\hat{c}(x)$。若 $\hat{c}(x) = c(x)$，则译码正确；否则，若 $\hat{c}(x) \neq c(x)$，则译码错误。

译码器实现的复杂程度，往往是一个纠错码能否实用的关键。伴随式计算电路比较容易实现，困难的是根据伴随式 $s(x)$找出对应的估值错误图样 $\hat{e}(x)$。对于（n，k）循环码来说，伴随式长度有（$n-k$）位，有 2^{n-k} 种不同的伴随式，当 n，k 很大时，这种查找变得非常困难。利用

循环码的循环特性，经常会使其译码运算变得简单，这也是循环码受到关注和重视的重要原因。至今，对于循环码人们已经找到了许多有效的译码方法，大致有代数译码和概率译码两类。本节主要介绍循环码译码的基本概念。

根据定理 9.6 可知，在计算得到接收码字多项式 $y(x)$的伴随式 $s(x)$后，无须重新计算就可得到 $y(x)$的各次循环移位所对应的伴随式 $s^{(i)}(x)$，$i=1$，2，…，$n-1$。由式（9-21）可知，若 $s(x)$是 $y(x)$的伴随式，则它也是 $e(x)$的伴随式，而 $s(x)$在伴随式计算电路中自发运算所得的 $s^{(i)}(x)$也就是 $e(x)$的各次循环移位所对应的伴随式。这样就可以把某一可纠正的错误图样 $e(x)$及其所有的小于等于 $n-1$ 次的循环移位归成一类，只需用一个错误图样来代表。译码时只要计算这个错误图样的伴随式，该类中其他错误图样的伴随式都可由该伴随式在 $g(x)$除法电路中循环移位来得到。这样做的好处是使译码器要识别的错误图样的个数大为减少，从而有望降低译码器的复杂程度。

【例 9.12】 仍以例 9.11 中的（7，3）循环码为例。当码字传送出现一位错误时，若用一般译码器，需要识别（0 0 0 0 0 0 1），（0 0 0 0 0 1 0），（0 0 0 0 1 0 0），（0 0 0 1 0 0 0），（0 0 1 0 0 0 0），（0 1 0 0 0 0 0），（1 0 0 0 0 0 0）7 个不同的错误图样，但对于按定理 9.6 设计的循环码译码器来说，可以把这些错误图样归成一类，译码器只要识别其中的一个错误图样就可以了。下面就来设计这样的译码器。

（7，3）循环码的生成矩阵为 $g(x)=x^4+x^3+x^2+1$，其译码器电路示于图 9-7。

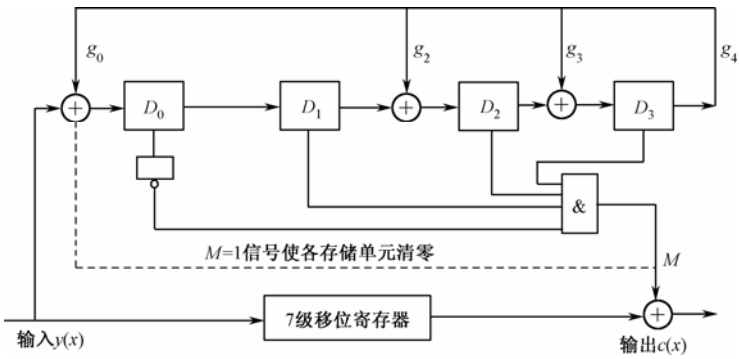


图 9-7 （7，3）循环码译码器

设发送码字 $\boldsymbol{c}$=（0 1 1 1 0 1 0），错误图样 $\boldsymbol{e}$=（1 0 0 0 0 0 0），则接收矢量 $\boldsymbol{y}$=（1 1 1 1 0 1 0），伴随式 $s(x)=e(x)\ [\mathrm{mod}\ g(x)]=x^6\ [\mathrm{mod}\ (x^4+x^3+x^2+1)]=x^3+x^2+x$，即 $\boldsymbol{s}$=（1 1 1 0）。

译码器工作过程如表 9-11 所示。清零后在第 1～7 拍，接收码字 $\boldsymbol{y}$ 一方面被送入移位寄存器 D_0～D_3，另一方面被送入 $g(x)$除法电路计算伴随式，在第 7 拍结束时得到 $\boldsymbol{s}$=（1 1 1 0）。随后与门开，并输出纠错信号 $M=1$，对接收码字 $\boldsymbol{y}$ 的第一位实施纠错，同时 $M=1$ 信号使 D_0～D_3 清零。第 8～14 拍移位寄存器中的内容依次输出，由于错误码元已被纠正，故输出 $\hat{\boldsymbol{c}}$=（0 1 1 1 0 1 0）。

若上述码字传送时，错误图样是 $\boldsymbol{e}$=（0 0 0 1 0 0 0），则接收矢量 $\boldsymbol{y}$=（0 1 1 0 0 1 0）。译码器工作过程如表 9-12 所示。从表中可见，到第 7 拍结束时 $\boldsymbol{s}$ 不是全 0，但也不是（1 1 1 0），说明接收码字有错，但错误图样不是（1 0 0 0 0 0 0）。从第 8 拍开始 $g(x)$除法电路进行自发运算，到第 10 拍结束时，得 $\boldsymbol{s}$=（1 1 1 0）。然后与门开并输出纠错信号 $M=1$，对 $\boldsymbol{y}$ 的对应位实施纠错，同时 $M=1$ 信号使 D_0～D_3 清零。

表 9-11 （7，3）循环码译码过程 1

节 拍	y	M	移位寄存器状态							伴随式状态				$\hat{c}$
			x^0	x^1	x^2	x^3	x^4	x^5	x^6	D_0	D_1	D_2	D_3	
0			0	0	0	0	0	0	0	0	0	0	0	
1	1	0	1	0	0	0	0	0	0	1	0	0	0	
2	1	0	1	1	0	0	0	0	0	1	1	0	0	
3	1	0	1	1	1	0	0	0	0	1	1	1	0	
4	1	0	1	1	1	1	0	0	0	1	1	1	1	
5	0	0	0	1	1	1	1	0	0	1	1	0	0	
6	1	0	1	0	1	1	1	1	0	1	1	1	0	
7	0	0	0	1	0	1	1	1	1	0	1	1	1	
8		1	0	0	1	0	1	1	1	0	0	0	0	0
9		0	0	0	0	1	0	1	1	0	0	0	0	1
10		0	0	0	0	0	1	0	1	0	0	0	0	1
11		0	0	0	0	0	0	1	0	0	0	0	0	1
12		0	0	0	0	0	0	0	1	0	0	0	0	0
13		0	0	0	0	0	0	0	0	0	0	0	0	1
14		0	0	0	0	0	0	0	0	0	0	0	0	0

表 9-12 （7，3）循环码译码过程 2

节 拍	y	M	移位寄存器状态							伴随式状态				$\hat{c}$
			x^0	x^1	x^2	x^3	x^4	x^5	x^6	D_0	D_1	D_2	D_3	
0			0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	1	0	0	0	0	0	0	1	0	0	0	
3	1	0	1	1	0	0	0	0	0	1	1	0	0	
4	0	0	0	1	1	0	0	0	0	0	1	1	0	
5	0	0	0	0	1	1	0	0	0	0	0	1	1	
6	1	0	1	0	0	1	1	0	0	0	0	1	0	
7	0	0	0	1	0	0	1	1	0	0	0	0	1	
8		0	0	0	1	0	0	1	1	1	0	1	1	0
9		0	0	0	0	1	0	0	1	1	1	1	0	1
10		0	0	0	0	0	1	0	0	0	1	1	1	1
11		1	0	0	0	0	0	1	0	0	0	0	0	1
12		0	0	0	0	0	0	0	1	0	0	0	0	0
13		0	0	0	0	0	0	0	0	0	0	0	0	1
14		0	0	0	0	0	0	0	0	0	0	0	0	0

9.5.3 Meggit译码器

9.5.2 节所述的循环码译码器的缺点是，译一个码字需要 $2n$ 节拍，因而无法对接收到的码字实现连续的译码输出。改进的译码器称为 Meggit 通用译码器，其结构如图 9-8 所示。将 $s(x)$ 的计算电路与 $s(x)$的自发运算电路并行完成，从而可实现连续的译码输出。

接收码字 $y(x)$一方面被送入 n 级移位寄存器，一方面被送入 $s(x)$计算电路。经 n 节拍后，将在 $s(x)$计算电路中得到的 $s(x)$送入自发运算电路。自发运算电路在结构上与 $s(x)$计算电路相

同。从 n+1 拍至 $2n$ 拍完成该码字对应伴随式的自发运算及纠错、译码输出。与此同时，第二个接收码字一方面被送入 n 级移位寄存器，一方面被送入 $s(x)$计算电路。可见，前一个码字的纠错译码过程与后一个码字的 $s(x)$计算过程在时间上是重叠的。虽然每一个码字在译码器中仍需逗留 $2n$ 拍，但从整体上看，该译码器实现了连续的译码输出。

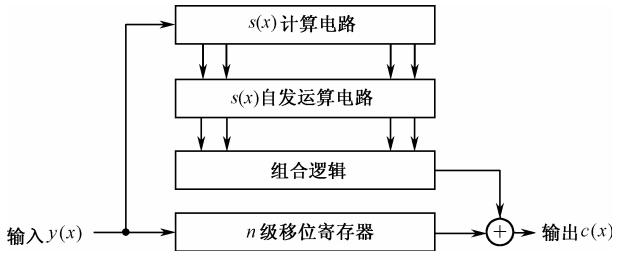


图 9-8　Meggit 通用译码器

【例 9.13】　GF(2)中的（7，4）循环码，生成矩阵 $g(x) = x^3+x+1$，利用 Meggit 通用译码器译码，电路如图 9-9 所示，图中 S_0, S_1, S_2 代表移位寄存器，由它们构成伴随式寄存器。

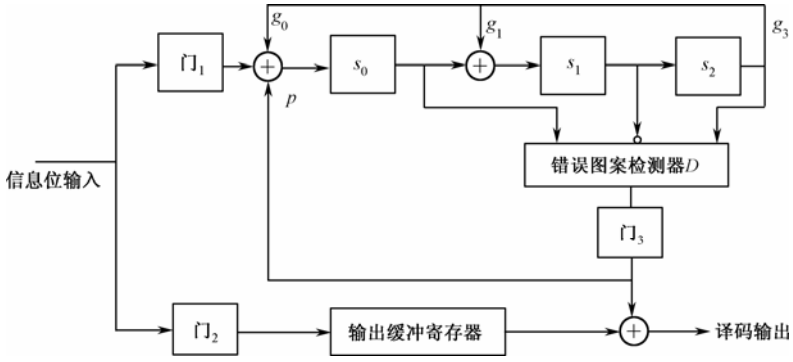


图 9-9　（7，4）循环码 Meggit 译码电路

译码过程可叙述如下。

先考虑接收矢量最高位 y_6：如果 y_6 出错，则错误图案为 $e(x) = x_6$，对应伴随式 $s^{(6)}(x) = x^6[\mathrm{mod}\, g(x)] = x^2+1$。

如果其他位出错，类似地也可算得对应的伴随式，把各种错误图案所对应的伴随式 $s^{(i)}(x)$（i= 0,1,2,3,4,5,6）列于表 9-13。

表 9-13　各种错误图案所对应的伴随式

错 误 图 案	伴　随　式	s_0 s_1 s_2
x^6	x^2+1	1　0　1
x^5	x^2+x+1	1　1　1
x^4	x^2+x	0　1　1
x^3	x+1	1　1　0
x^2	x^2	0　0　1
x	x	0　1　0
1	1	1　0　0

译码步骤：

（1）门 1 开，门 2 开，门 3 关，接收矢量 $\boldsymbol{y}$ 通过门 1 依次进入 s_0，s_1，s_2，并通过门 2 进入输出缓冲器。当第 7 位数字送完，门 1 关，s_0，s_1，s_2 中就是伴随式 $s(x)$的系数，若其内容与 $s^{(i)}(x)$对应，就表示第 i（$i=0, 1, \cdots, 6$）位出错；

（2）当 s_0，s_1，$s_2=101$ 时，说明最高位 y_6 有错，此时错误图案检测器 D 输出 1。门 3 开，信号 $p=1$ 一方面纠正最高位，一方面反馈至伴随式寄存器，使伴随式寄存器反馈右移成 000，由于门 1 关，门 3 开，故此后 D 的输出通过反馈线就充当了输入的作用，由于每次都输入为 0（只有 s_0，s_1，$s_2=101$ 才能使 $D=1$），所以伴随式每次都是 000，后面的各位输出不再纠正，直至全部接收矢量输出完毕；

（3）当 s_0，s_1，$s_2=111$ 时，说明第 5 位 y_5 有错，此时错误图案检测器 D 输出 0。门 3 开，按原值输出最高位 y_6，其余 y_0，…，y_5 右移一位，将 y_5 移在最高位。同时错误图案检测器 D 输出的 $p=0$ 信号反馈至伴随器移位寄存器，使伴随器右移反馈成 101。这时 $D=1$，一方面纠正 y_5，另一方面反馈至伴随器寄存器，使伴随器寄存器反馈右移成 000，此后 s_0，s_1，s_2 一直为 000，直至全部接收矢量输出完毕；

（4）若 s_0，s_1，s_2 等于其他值，从上面的叙述可知，Meggit 译码器可纠正 s_0，s_1，s_2 所对应的所有错误图案，且该（7，3）码只能纠一位错。

虽然 Meggit 通用译码器提供了循环码译码的一般方法，但在具体实现时，随着码长的增加和纠错能力的提高，其组合逻辑电路部分会变得很复杂，甚至难以实现。人们为解决循环码的译码问题进行了大量的研究，至今已提出了多种各有特色的循环码译码方法，如捕错译码、大数逻辑译码以及用于 BCH 码的迭代译码和快速译码等，这些研究极大地推动了循环码的广泛应用。

9.6 一些重要的循环码

我们已经研究了循环码的特点及编译码原理，本节将介绍一些重要的循环码，如循环 Hamming 码、BCH 码等，它们已得到了广泛的应用。

9.6.1 循环Hamming码

我们已经知道，循环码的码字必是生成多项式 $g(x)$的倍式。那么，如果α是生成多项式 $g(x)$的根，则α一定也是码字多项式 $c(x)$的根。既然循环码可以用 $g(x)$来定义，那么我们也可以用 $g(x)$的根α 来定义它。

定义 9.2 设α是 GF(2^m)上的一个本原元，则以α的本原多项式为生成多项式的（2^m-1，2^m-1-m）Hamming 码是循环码。

根据定义，如果

$$\boldsymbol{c}=(c_{n-1}\ c_{n-2}\ \cdots\ c_1\ c_0)$$

是 Hamming 循环码的一个码字，而α是生成多项式 $g(x)$的根，则

$$c(\alpha)=0$$

或

$$c_{n-1}\alpha^{n-1}+c_{n-2}\alpha^{n-2}+\cdots+c_1\alpha+c_0=0$$

上式写成矩阵形式为

$$\left[\alpha^{n-1} \quad \alpha^{n-2} \quad \cdots \quad \alpha \quad 1\right]\cdot\left[c_{n-1} \quad c_{n-2} \quad \cdots \quad c_1 \quad c_0\right]^T=0 \tag{9-22}$$

由此可得码的校验矩阵为

$$\boldsymbol{H}=\left[\alpha^{n-1} \quad \alpha^{n-2} \quad \cdots \quad \alpha \quad 1\right] \tag{9-23}$$

因码长 $n=2^m-1$，故 $\boldsymbol{H}$也可以表示为

$$\boldsymbol{H}=\left[\alpha^{2^m-2} \quad \alpha^{2^m-3} \quad \cdots \quad \alpha \quad 1\right] \tag{9-24}$$

例如，以 GF(2^3)上的三次本原多项式为生成多项式，可生成一个（7，4）循环 Hamming 码，其生成多项式 $g(x)=x^3+x+1$。

设α是 GF(2^3)上的本原元，则码的校验矩阵

$$\boldsymbol{H}=\left[\alpha^6 \quad \alpha^5 \quad \alpha^4 \quad \alpha^3 \quad \alpha^2 \quad \alpha^1 \quad \alpha^0\right]=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{bmatrix}$$

又如，以 GF(2^4)上的四次本原多项式为生成多项式，可生成一个（15，11）循环 Hamming 码，其生成多项式 $g(x)=x^4+x+1$。码的校验矩阵为

$$\boldsymbol{H}=\left[\alpha^{14} \quad \alpha^{13} \quad \cdots \quad \alpha \quad 1\right]=\begin{bmatrix}1&1&1&1&0&1&0&1&1&0&0&1&0&0&0\\0&1&1&1&1&0&1&0&1&1&0&0&1&0&0\\0&0&1&1&1&1&0&1&0&1&1&0&0&1&0\\1&1&1&0&1&0&1&1&0&0&1&0&0&0&1\end{bmatrix}$$

9.6.2 BCH码

由 A.Hocquenghem，R.C.Bose 和 D.K.Ray-Chaudhuri 分别提出的 BCH 码是一类可纠正多个随机错误的循环码，也是 Hamming 码在纠多重错方面的重要推广。BCH 码是目前线性分组码中最有效的一种码，具有很强的纠错能力。这里仅讨论二元的本原 BCH 码。

循环 Hamming 码是纠单个错误的完备码，其码字中每一位差错所对应的伴随式 $S(x)$都恰好是 $\boldsymbol{H}$矩阵中的对应列；但如果码字发生两位错，则相应的伴随式 $S(x)$就会是 $\boldsymbol{H}$矩阵中对应两列之和，而此两列之和必等于 $\boldsymbol{H}$矩阵中的另外某一列，因而将会造成错误译码。要提高码的纠错能力，必须增加码的冗余度。BCH 码就是在 Hamming 码的基础上，通过增加 $\boldsymbol{H}$矩阵的行数来提高码的纠错能力的。

定义 9.3 设α是 GF(2^m)上的一个本原元，t 为整数，则以含有α，α^2，…，α^{2t} 等共 $2t$ 个根、其系数在 GF(2)上的最低次多项式 $g(x)$为生成多项式的循环码，称为二元本原 BCH 码。

由定义可知，二元本原 BCH 码的生成多项式 $g(x)$的全部根为α，α^2，…，α^{2t}及其共轭根组。令 $m_i(x)$为α^i 的最小多项式，则 $g(x)$必是 $m_1(x)$，$m_2(x)$，…，$m_{2t}(x)$的最小公倍式，

即 $$g(x)=[m_1(x)，m_2(x)，\cdots，m_{2t}(x)]$$

显然，在 GF(2^m)上，α^{2i}与α^i 具有相同的最小多项式，所以有

$$g(x)=[m_1(x)，m_3(x)，\cdots，m_{2t-1}(x)] \tag{9-25}$$

综上所述，二元本原 BCH 码的参数为

码长 $n=2^m-1$

校验位数 $r=n-k\leqslant mt$

最小距离 $d \geqslant 2t+1$

纠错能力为 t。

若α，α^3，…，α^{2t-1} 是二元本原 BCH 码的生成多项式 $g(x)$的根，则由于码字多项式 $c(x)$是生成多项式 $g(x)$的倍式，故α，α^3，…，α^{2t-1} 也必是码字多项式 $c(x)$的根，即

$$c_{n-1}(\alpha^i)^{n-1}+c_{n-2}(\alpha^i)^{n-2}+\cdots+c_1\alpha^i+c_0=0 \qquad i=1,3,\cdots,2t-1$$

写成矩阵形式为

$$\begin{bmatrix} \alpha^{n-1} & \alpha^{n-2} & \cdots & \alpha & 1 \\ (\alpha^3)^{n-1} & (\alpha^3)^{n-2} & \cdots & \alpha^3 & 1 \\ \vdots & & & & \\ (\alpha^{2t-1})^{n-1} & (\alpha^{2t-1})^{n-2} & \cdots & \alpha^{2t-1} & 1 \end{bmatrix} \cdot \begin{bmatrix} c_{n-1} \\ c_{n-2} \\ \vdots \\ c_0 \end{bmatrix} = \boldsymbol{0} \qquad (9\text{-}26)$$

故码的校验矩阵为

$$\boldsymbol{H} = \begin{bmatrix} \alpha^{n-1} & \alpha^{n-2} & \cdots & \alpha & 1 \\ (\alpha^3)^{n-1} & (\alpha^3)^{n-2} & \cdots & \alpha^3 & 1 \\ \vdots & & & & \\ (\alpha^{2t-1})^{n-1} & (\alpha^{2t-1})^{n-2} & \cdots & \alpha^{2t-1} & 1 \end{bmatrix} \qquad (9\text{-}27)$$

在 $\boldsymbol{H}$矩阵中，GF(2^m)上的每个元素α^i（$i=1$，3，…，$2t-1$），可用二进制 m 重表示，因此 $\boldsymbol{H}$矩阵至多只有 mt 行，说明码的校验元至多只有 mt 个。

【例 9.14】 设 $m=4$，α是 GF(2^4)上的本原元，求码长 $n=2^4-1=15$ 的二元本原 BCH 码。

若 $t=1$，则码以α为根，即以α，α^2，α^4，α^8 共轭根组为根，最小多项式

$$m_1(x)=x^4+x+1$$

生成多项式 $$g(x)=m_1(x)=x^4+x+1$$

校验位数目 $n-k=4$，由此生成的（15，11）BCH 码就是已经学过的循环 Hamming 码。

显然，纠单错的二元本原 BCH 码就是循环 Hamming 码。

若 $t=2$，则码以α，α^3 为根，即以α，α^2，α^4，α^8 共轭根组和α^3，α^6，α^{12}，$\alpha^{24}=\alpha^9$ 共轭根组为根，最小多项式

$$m_3(x)=x^4+x^3+x^2+x+1$$

生成多项式

$$\begin{aligned} g(x) &= m_1(x)m_3(x) \\ &= (x^4+x+1)(x^4+x^3+x^2+x+1) \\ &= x^8+x^7+x^6+x^4+1 \end{aligned}$$

校验位数目 $n-k=8$，由此生成的（15，7）BCH 码，其校验矩阵

$$\begin{aligned} \boldsymbol{H} &= \begin{bmatrix} \alpha^{14} & \alpha^{13} & \alpha^{12} & \alpha^{11} & \alpha^{10} & \alpha^9 & \alpha^8 & \alpha^7 & \alpha^6 & \alpha^5 & \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \\ \alpha^{42} & \alpha^{39} & \alpha^{36} & \alpha^{33} & \alpha^{30} & \alpha^{27} & \alpha^{24} & \alpha^{21} & \alpha^{18} & \alpha^{15} & \alpha^{12} & \alpha^9 & \alpha^6 & \alpha^3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \alpha^{14} & \alpha^{13} & \alpha^{12} & \alpha^{11} & \alpha^{10} & \alpha^9 & \alpha^8 & \alpha^7 & \alpha^6 & \alpha^5 & \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \\ \alpha^{12} & \alpha^9 & \alpha^6 & \alpha^3 & \alpha^0 & \alpha^{12} & \alpha^9 & \alpha^6 & \alpha^3 & \alpha^0 & \alpha^{12} & \alpha^9 & \alpha^6 & \alpha^3 & 1 \end{bmatrix} \end{aligned}$$

用二进制 4 重表示 $\boldsymbol{H}$矩阵中的α^i，得

$$\boldsymbol{H}=\begin{bmatrix}1&1&1&0&1&0&1&1&0&0&1&0&0&0&1\\0&0&1&1&1&1&0&1&0&1&1&0&0&1&0\\0&1&1&1&1&0&1&0&1&1&0&0&1&0&0\\1&1&1&1&0&1&0&1&1&0&0&1&0&0&0\\1&0&0&0&1&1&0&0&0&1&1&0&0&0&1\\1&1&0&0&0&1&1&0&0&0&1&1&0&0&0\\1&0&1&0&0&1&0&1&0&0&1&0&1&0&0\\1&1&1&1&0&1&1&1&1&0&1&1&1&1&0\end{bmatrix}$$

若 $t=3$，则码以α，α^3，α^5 为根，即以α，α^2，α^4，α^8 共轭根组，α^3，α^6，α^{12}，$\alpha^{24}=\alpha^9$ 共轭根组和α^5，α^{10} 共轭根组为根，最小多项式

$$m_5(x)=x^2+x+1$$

生成多项式

$$\begin{aligned}g(x)&=m_1(x)m_3(x)m_5(x)\\&=(x^4+x+1)(x^4+x^3+x^2+x+1)(x^2+x+1)\\&=x^{10}+x^8+x^5+x^4+x^2+x+1\end{aligned}$$

校验位数目 $n-k=10$，由此生成的（15，5）BCH 码，其校验矩阵

$$\boldsymbol{H}=\begin{bmatrix}\alpha^{14}&\alpha^{13}&\alpha^{12}&\alpha^{11}&\alpha^{10}&\alpha^{9}&\alpha^{8}&\alpha^{7}&\alpha^{6}&\alpha^{5}&\alpha^{4}&\alpha^{3}&\alpha^{2}&\alpha&1\\\alpha^{42}&\alpha^{39}&\alpha^{36}&\alpha^{33}&\alpha^{30}&\alpha^{27}&\alpha^{24}&\alpha^{21}&\alpha^{18}&\alpha^{15}&\alpha^{12}&\alpha^{9}&\alpha^{6}&\alpha^{3}&1\\\alpha^{70}&\alpha^{65}&\alpha^{60}&\alpha^{55}&\alpha^{50}&\alpha^{45}&\alpha^{40}&\alpha^{35}&\alpha^{30}&\alpha^{25}&\alpha^{20}&\alpha^{15}&\alpha^{10}&\alpha^{5}&1\end{bmatrix}$$

$$=\begin{bmatrix}\alpha^{14}&\alpha^{13}&\alpha^{12}&\alpha^{11}&\alpha^{10}&\alpha^{9}&\alpha^{8}&\alpha^{7}&\alpha^{6}&\alpha^{5}&\alpha^{4}&\alpha^{3}&\alpha^{2}&\alpha&1\\\alpha^{12}&\alpha^{9}&\alpha^{6}&\alpha^{3}&\alpha^{0}&\alpha^{12}&\alpha^{9}&\alpha^{6}&\alpha^{3}&\alpha^{0}&\alpha^{12}&\alpha^{9}&\alpha^{6}&\alpha^{3}&1\\\alpha^{10}&\alpha^{5}&\alpha^{0}&\alpha^{10}&\alpha^{5}&\alpha^{0}&\alpha^{10}&\alpha^{5}&\alpha^{0}&\alpha^{10}&\alpha^{5}&\alpha^{0}&\alpha^{10}&\alpha^{5}&1\end{bmatrix}$$

$$=\begin{bmatrix}1&1&1&0&1&0&1&1&0&0&1&0&0&0&1\\0&0&1&1&1&1&0&1&0&1&1&0&0&1&0\\0&1&1&1&1&0&1&0&1&1&0&0&1&0&0\\1&1&1&1&0&1&0&1&1&0&0&1&0&0&0\\1&0&0&0&1&1&0&0&0&1&1&0&0&0&1\\1&1&0&0&0&1&1&0&0&0&1&1&0&0&0\\1&0&1&0&0&1&0&1&0&0&1&0&1&0&0\\1&1&1&1&0&1&1&1&1&0&1&1&1&1&0\\1&0&1&1&0&1&1&0&1&1&0&1&1&0&1\\1&1&0&1&1&0&1&1&0&1&1&0&1&1&0\\1&1&0&1&1&0&1&1&0&1&1&0&1&1&0\\0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\end{bmatrix}$$

由于矩阵中第 10 行和第 11 行相关，第 12 行为全 0，故可删去第 11 行和第 12 行，最终得到（15，5）BCH 码的校验矩阵是一个 10 行 15 列的矩阵，即

$$\boldsymbol{H}=\begin{bmatrix}1&1&1&0&1&0&1&1&0&0&1&0&0&0&1\\0&0&1&1&1&1&0&1&0&1&1&0&0&1&0\\0&1&1&1&1&0&1&0&1&1&0&0&1&0&0\\1&1&1&1&0&1&0&1&1&0&0&1&0&0&0\\1&0&0&0&1&1&0&0&0&1&1&0&0&0&1\\1&1&0&0&0&1&1&0&0&0&1&1&0&0&0\\1&0&1&0&0&1&0&1&0&0&1&0&1&0&0\\1&1&1&1&0&1&1&1&1&0&1&1&1&1&0\\1&0&1&1&0&1&1&0&1&1&0&1&1&0&1\\1&1&0&1&1&0&1&1&0&1&1&0&1&1&0\end{bmatrix}$$

本 章 小 结

循环码是线性分组码中一个重要的子类，由于这种码的代数结构完全建立在有限域基础上，它是目前理论上最为成熟的一类码。本章的主要内容如下。

（1）循环码的定义及多项式描述：循环码的循环特性、循环码的码字多项式；

（2）循环码的性质：循环码的生成多项式、生成多项式的重要性质、由生成多项式构造生成矩阵；循环码的校验多项式，由校验多项式构造校验矩阵；

（3）循环码的编码：除法电路，用生成多项式 $g(x)$编码的理论及实现电路，用校验多项式 $h(x)$编码的理论及实现电路；

（4）循环码的译码：伴随式的计算电路、自发运算电路、Meggit 译码器；

（5）循环 Hamming 码和 BCH 码：用 $g(x)$的根定义循环码，建立在有限域扩域上的 BCH 码。

本章的重点是循环码的代数理论及其编译码方法，应结合实例搞清概念。

思考题与习题

9.1　什么是循环码？如何用多项式来描述一个循环码？

9.2　循环码的生成多项式是如何定义的？生成多项式 $g(x)$有什么特点和性质？

9.3　循环码的生成多项式 $g(x)$和校验多项式 $h(x)$之间有什么关系？如何在已知码的生成多项式和校验多项式时，得到对应的生成矩阵和校验矩阵？

9.4　试述利用生成多项式 $g(x)$实现循环码编码的步骤，如何用电路实现编码？

9.5　试述利用校验多项式 $h(x)$实现循环码编码的步骤，如何用电路实现编码？

9.6　二元（15，8）循环码共有多少码字？能否由一个码字的循环产生所有的码字？为什么？

9.7　利用接收序列 $y(x)$的伴随式 $S(x)$进行检错的原理是什么？

9.8　Meggit 通用译码器有什么特点？为什么这种译码器能够实现连续的译码输出？

9.9　设循环码由 $g(x)=x^8+x^7+x^6+x^4+1$ 生成，证明此码长度为 15。

9.10　已知（7，3）循环码的生成多项式 $g(x)=x^4+x^2+x+1$，求其校验多项式 $h(x)$。

9.11　$x^{15}+1$ 在 GF（2）上可分解为以下既约多项式的乘积：

$$x^{15}+1=(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$$

当构成（15，9）码时，有多少种不同的选择？分别写出对应的生成多项式及校验多项式。

9.12　已知（15，7）循环码的生成多项式 $g(x)=x^8+x^7+x^6+x^4+1$。

（1）求出该循环码的生成矩阵和校验矩阵，并变换为标准形式；

（2）问：$y(x)=x^7+x^5+x^3+x+1$ 是码多项式吗？

（3）求出 $y(x)$的伴随式。

9.13　设计一个由 $g(x)=x^4+x+1$ 生成的（15，11）循环汉明码的编码器。

9.14　证明：$x^{10}+x^8+x^5+x^4+x^2+x+1$ 为（15，5）循环码的生成多项式，并求：

（1）该码的生成矩阵；

（2）当信息多项式为 $m(x)=x^4+x^2+x+1$ 时的系统码多项式；

（3）画出以 $g(x)$除法电路为核心的 $n-k$ 级编码器。

9.15　证明：由 $g(x)=x^{10}+x^7+x^6+x^4+x^2+1$ 可生成一个（21，11）循环码。画出此码的伴随式计算电路。若接收码字多项式为 $y(x)=x^{17}+x^5+1$，其伴随式是什么？

9.16　已知一个（n，k）循环码的生成多项式为 $g(x)$，$x+1$ 是 $g(x)$的一个因子。求证：

（1）设 n 为奇数，则全 1 的 n 重矢量不是一个码字；

（2）设 n 为偶数，则全 1 的 n 重矢量是一个码字。

9.17　设计一个由 $g(x)=x^4+x+1$ 生成的（15，11）循环汉明码的译码器。

9.18　若 $g(x)$生成二元（n，k）循环码，证明其反多项式 $g^*(x)=x^{n-1}g(x)$也能生成二元（n，k）循环码。

9.19　已知 $x^{15}-1=(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$，试问有多少个长为 15 的循环码？给出 4 个这类码及其对偶码的生成多项式，并指出这 8 个码的维数。

9.20　证明：以下述 GF(2) 的 5×8 矩阵为生成矩阵的码是循环码，并给出其生成多项式和一致校验多项式。

$$\boldsymbol{G}=\begin{bmatrix}1&0&0&0&0&1&1&1\\0&1&0&0&0&1&0&0\\0&0&1&0&0&0&1&0\\0&0&0&1&0&0&0&1\\0&0&0&0&1&1&1&1\end{bmatrix}$$

第 10 章
卷 积 码

内容提要

卷积码是 1955 年由 Elias 提出的，它是一种非常重要的差错控制编码。卷积码编码时，所编码的码段不仅与当前的信息段有关，而且与前面若干时刻输入至编码器的信息段有关。同样，卷积码译码时也要结合当前时刻和以前各时刻接收到的码段来提取有关信息。由于卷积码充分利用了前后码段之间的相关性，故与分组码相比较，卷积码的性能更好，译码更容易。本章首先介绍卷积码的基本概念，然后介绍卷积码的数学描述方法和图形描述方法，最后介绍一种最佳的卷积码概率译码方法——Viterbi 译码。

知识要点

卷积码的概念，卷积码的矩阵和多项式描述，树图和网格图描述，Viterbi 译码方法。

教学建议

卷积码和线性分组码有许多相似之处，建议结合实例讲解，学时数为 6 学时。

10.1 卷积码基本概念

前面介绍的（n，k）分组码，无论编码还是译码，一个码组中的 $n-k$ 个检验元仅与本组的 k 个信息元有关，而与其他各组码元无关。从这个意义上来说分组码是无记忆的。卷积码是有记忆的，它用（n，k，m）表示，卷积码编码时 n–k 个检验元不仅与当前时刻的 k 个信息元有关，而且与前面 m 个时刻的信息段有关，因此卷积码的编码器中需要有存储 m 个信息段的记忆部件。定义 m 为编码存储，它代表了信息段需存储的级数。定义（$m+1$）为编码约束度，表明编码过程中互相约束的码段个数。与分组码一样，码率为 $\frac{k}{n}$，由于卷积码前后码段之间具有相关性，故 k 和 n 一般较小，码率也较低。由于码率低，故卷积码的冗余度高，纠错能力较强。

卷积码编码器的原理图如图 10-1 所示。

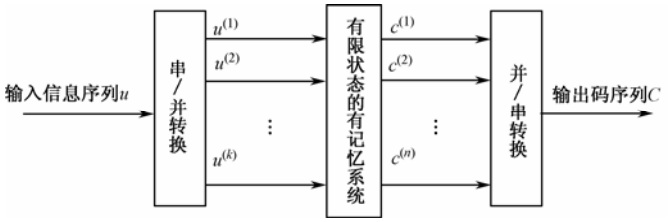


图 10-1　卷积码编码器原理图

为了更好地阐述卷积码的基本概念，下面以二进制卷积码编码器为例进行描述。

【例 10.1】　图 10-2 所示为（2，1，2）卷积码编码器，分析其工作原理。

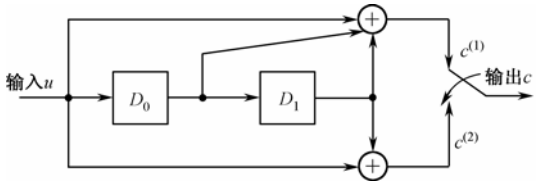


图 10-2　（2，1，2）卷积码编码器

D_0，D_1 为两级移位寄存器，信息序列 $\boldsymbol{u}$ 在时钟节拍下串行输入，编码器输出端由开关实现并/串转换，每一个时间单位（n 个码元）获得一个码段送入信道。图中码段长 $n=2$，每个码段的信息位长度 $k=1$，编码存储 $m=2$。

由图可列出表达式：

$$\boldsymbol{u}=[u_0\ u_1u_2\cdots] \qquad \boldsymbol{c}=[c_0^{(1)}c_0^{(2)}c_1^{(1)}c_1^{(2)}c_2^{(1)}c_2^{(2)}\cdots]$$

$$c_i^{(1)}=u_{i-2}+u_{i-1}+u_i \qquad \boldsymbol{c}^{(1)}=c_0^{(1)}c_1^{(1)}c_2^{(1)}\cdots$$

$$c_i^{(2)}=u_{i-2}+u_i \qquad \boldsymbol{c}^{(2)}=c_0^{(2)}c_1^{(2)}c_2^{(2)}\cdots$$

式中，下标表示时刻，数值小者时间上在前，可见第 i 时刻的码段 $\boldsymbol{c}_i=c_i^{(1)}c_i^{(2)}$ 不仅与当前的信息位 u_i 有关，还与前 2 个码段的信息位 u_{i-1}，u_{i-2} 有关。假设移位寄存器 D_0D_1 初始状态为全 0，则当输入信息序列 $\boldsymbol{u}=[100]$ 时，编码器的工作过程如表 10-1 所示。

表 10-1　（2，1，2）码编码过程

时 钟 节 拍	输入 u	移存器初态	移存器次态	输出 $\boldsymbol{c}$
1	1	00	10	11
2	0	10	01	10
3	0	01	00	11

所以输入信息序列 $\boldsymbol{u}$ = [100] 时，输出码字 $\boldsymbol{c}$ = [11 10 11]。观察发现每个码段的最左边码元与信息码元不同，所以图 10-2 是非系统码编码器。在同样的参数条件下，系统卷积码与非系统卷积码的抗干扰性能不一定相同。下面再介绍一个系统码的例子。

【例 10.2】　图 10-3 是（3，2，2）系统卷积码编码器，分析输入 $\boldsymbol{u}$ = [10 00 00] 和 $\boldsymbol{u}$ = [01 00 00]时的输出码序列。

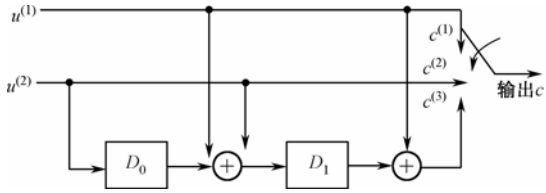


图 10-3　（3，2，2）卷积码编码器

第 i 时刻信息段 $\boldsymbol{u}_i = u_i^{(1)}u_i^{(2)}$，码段 $\boldsymbol{c}_i = c_i^{(1)}c_i^{(2)}c_i^{(3)}$，编码器由 2 个移存器构成，所以是（3，2，2）码。输入 $\boldsymbol{u}$ = [10 00 00] 的编码过程如表 10-2 所示，输出码 $\boldsymbol{c}$ = [101 001 000]。

表 10-2　（3，2，2）码编码过程

时 钟 节 拍	输入 u_i	移位寄存器初态	移位寄存器次态	输出 c_i
1	10	00	01	101
2	00	01	00	001
3	00	00	00	000

同理，可得输入 $\boldsymbol{u}$ = [01 00 00] 的编码输出 $\boldsymbol{c}$ =[010 001 001]。可以看出码段的左边 2 个码元和输入的 2 个信息元始终一致，所以是系统码。

10.2　卷积码的数学描述

描述卷积码编、译码过程的方法很多，大致分为两种类型：解析表示法和图形表示法。在解析表示法中又可分为卷积方法、矩阵方法和多项式方法，图形表示法也可分为状态图法、树图法和网格图法。卷积码的表示方法与它所采用的译码方法有密切关系。本节只介绍与代数译码相关的矩阵和多项式表示方法。

10.2.1　卷积码的矩阵描述

在卷积码中，无论是系统码还是非系统码，码的生成矩阵一旦确定，则码也就完全确定了。

【例 10.3】　以图 10-2 的（2，1，2）非系统卷积码为例来讨论生成矩阵。

$$c_0^{(1)}=u_0 \qquad c_0^{(2)}=u_0$$
$$c_1^{(1)}=u_0+u_1 \qquad c_1^{(2)}=u_1$$
$$c_2^{(1)}=u_0+u_1+u_2 \qquad c_2^{(2)}=u_0+u_2$$
$$c_3^{(1)}=u_1+u_2+u_3 \qquad c_3^{(2)}=u_1+u_3$$
$$\cdots$$

将上述方程组表示为矩阵，可得

$$\boldsymbol{C}=[c_0^{(1)}c_0^{(2)}c_1^{(1)}c_1^{(2)}c_2^{(1)}c_2^{(2)}\cdots]=[u_0u_1u_2u_3\cdots]\cdot\begin{bmatrix}1&1&1&0&1&1&0&0&\cdots\\0&0&1&1&1&0&1&1&\cdots\\0&0&0&0&1&1&1&0&\cdots\\0&0&0&0&0&0&1&1&\cdots\\&&&&\vdots&&&&\end{bmatrix}$$

即
$$\boldsymbol{C}=\boldsymbol{U}\boldsymbol{G}_\infty$$

$\boldsymbol{G}_\infty$被称为（2，1，2）卷积码的生成矩阵，这是一个半无限矩阵，重写如下：

$$\boldsymbol{G}_\infty=\begin{bmatrix}11&10&11&&&&\\&11&10&11&&0&\\&&11&10&11&&\\&0&&11&10&11&\\&&&&&&\ddots\end{bmatrix}$$

仔细观察 $\boldsymbol{G}_\infty$可发现它的每一行都是前一行右移 n 位的结果，也就是说它完全是由矩阵的第一行确定的。将第一行取出并表示为

$$\boldsymbol{g}_\infty=[11\ 10\ 11\ 00\ \cdots]$$

称 $\boldsymbol{g}_\infty$ 为该码的基本生成矩阵，通过与表 10-1 比较，$\boldsymbol{g}_\infty$ 其实就是当 $\boldsymbol{u}=[100\cdots]$，即输入信息序列为冲激序列时卷积码编码器的冲激响应。

令信息序列 $\boldsymbol{u}=[10101]$，则输出码字

$$\boldsymbol{C}=\boldsymbol{U}\cdot\boldsymbol{G}_\infty=[10101]\begin{bmatrix}11&10&11&&&&&\\&11&10&11&&0&&\\&&11&10&11&&&\cdots\\&0&&11&10&11&&\\&&&&11&10&11&\end{bmatrix}$$

$$=[11\ 10\ 00\ 10\ 00\ 10\ 11\ 00\ \cdots]$$

再如图 10-3 所示的系统卷积码，相应的冲激响应为 $\boldsymbol{u}=[10\ 00\ 00\ \cdots]$时，$\boldsymbol{c}=[101\ 001\ 000\ 000\cdots]$；$\boldsymbol{u}=[01\ 00\ 00\ \cdots]$时，$\boldsymbol{c}=[010\ 001\ 001\ 000\cdots]$

由 $\boldsymbol{u}=[10\ 00\ 00\ \cdots]$和 $\boldsymbol{u}=[01\ 00\ 00\ \cdots]$的冲激响应，得该码的基本生成矩阵为

$$\boldsymbol{g}_\infty=\begin{bmatrix}101&001&000&000&\cdots\\010&001&001&000&\cdots\end{bmatrix}$$

将 $\boldsymbol{g}_\infty$ 作为生成矩阵 $\boldsymbol{G}_\infty$ 的最上面两行，并经移位得到该码的生成矩阵为

$$\boldsymbol{G}_\infty=\begin{bmatrix}
101 & 001 & 000 & & & & \\
010 & 001 & 001 & & & & \\
 & 101 & 001 & 000 & & 0 & \\
 & 010 & 001 & 001 & & & \\
 & & 101 & 001 & 000 & & \\
 & & 010 & 001 & 001 & & \\
 & 0 & & 101 & 001 & 000 & \\
 & & & 010 & 001 & 001 & \\
 & & & & & & \ddots
\end{bmatrix}$$

显然，若输入信息序列 $\boldsymbol{u}=[10\ 11\ 01\ 11\ \cdots]$，则相应的输出码序列应为

$$\boldsymbol{C}=\boldsymbol{U}\cdot\boldsymbol{G}_\infty=\begin{bmatrix}10 & 11 & 01 & 11 & \cdots\end{bmatrix}\cdot\begin{bmatrix}
101 & 001 & 000 & 000 & 000 & 000 & \\
010 & 001 & 001 & 000 & 000 & 000 & \\
000 & 101 & 001 & 000 & 000 & 000 & \\
000 & 010 & 001 & 001 & 000 & 000 & \\
000 & 000 & 101 & 001 & 000 & 000 & \cdots \\
000 & 000 & 010 & 001 & 001 & 000 & \\
000 & 000 & 000 & 101 & 001 & 000 & \\
000 & 000 & 000 & 010 & 001 & 001 & \\
 & & & \vdots & & &
\end{bmatrix}$$

$$=[101\ 110\ 010\ 111\ 001\ 001\cdots]$$

一般情况下，（n，k，m）卷积码的生成矩阵可表示为

$$\boldsymbol{G}_\infty=\begin{bmatrix}
\boldsymbol{G}_0 & \boldsymbol{G}_1 & \boldsymbol{G}_2 & \cdots & \boldsymbol{G}_m & & \\
 & \boldsymbol{G}_0 & \boldsymbol{G}_1 & \cdots & \boldsymbol{G}_{m-1} & \boldsymbol{G}_m & 0 \\
 & & \boldsymbol{G}_0 & \cdots & \boldsymbol{G}_{m-2} & \boldsymbol{G}_{m-1} & \boldsymbol{G}_m \\
0 & & & & & & \ddots
\end{bmatrix}$$

基本生成矩阵 $\boldsymbol{g}_\infty=[\boldsymbol{G}_0\ \boldsymbol{G}_1\ \boldsymbol{G}_2\ \cdots\boldsymbol{G}_m\ 0\ \cdots]$

其中生成子矩阵 $\boldsymbol{G}_l=\begin{bmatrix}
g_{1,l}^{(1)} & g_{1,l}^{(2)} & \cdots & g_{1,l}^{(n)} \\
g_{2,l}^{(1)} & g_{2,l}^{(2)} & \cdots & g_{2,l}^{(n)} \\
\vdots & \vdots & & \vdots \\
g_{k,l}^{(1)} & g_{k,l}^{(2)} & \cdots & g_{k,l}^{(n)}
\end{bmatrix}$

生成矩阵中每一行的分组数（即码段数）为编码约束度 $m+1$，矩阵的总行数取决于输入信息序列的长度。

10.2.2 卷积码的多项式描述

通过延时算子 x 表示卷积码编码过程中一个时间单位的延时，可以把输入输出序列用 x 的多项式表示。

例如，第 i 路输入信息序列 $\boldsymbol{u}^{(i)}=[u_0^{(i)}u_1^{(i)}u_2^{(i)}\cdots]$ 和第 j 路输出码序列 $\boldsymbol{C}^{(j)}=[c_0^{(j)}c_1^{(j)}c_2^{(j)}\cdots]$ 可分别表示为多项式

$$u^{(i)}(x)=u_0^{(i)}+u_1^{(i)}x+u_2^{(i)}x^2+\cdots$$

$$c^{(j)}(x)=c_0^{(j)}+c_1^{(j)}x+c_2^{(j)}x^2+\cdots$$

同样生成矩阵也可以表示成多项式形式，将生成子矩阵 $\boldsymbol{G}_l$（$0\leqslant l\leqslant m$）的 i 行 j 列取出组合后得到生成序列 $\boldsymbol{g}_i^{(j)}=[g_{i,0}^{(j)}g_{i,1}^{(j)}g_{i,2}^{(j)}\cdots g_{i,m}^{(j)}]$，$1\leqslant j\leqslant n$，$1\leqslant i\leqslant k$，表示成多项式为

$$g_i^{(j)}(x)=g_{i,0}^{(j)}+g_{i,1}^{(j)}x+\cdots+g_{i,m}^{(j)}x^m$$

$g_i^{(j)}(x)$ 表示 $\boldsymbol{u}^{(i)}\sim\boldsymbol{C}^{(j)}$的生成情况，其中 $g_{i,l}^{(j)}=1$代表了编码器中 $\boldsymbol{C}^{(j)}$生成时 $\boldsymbol{u}^{(i)}$的 l 次移位参与了异或，因此第 j 路输出的码多项式为

$$c^{(j)}(x)=u^{(1)}(x)g_1^{(j)}(x)+u^{(2)}(x)g_2^{(j)}(x)+\cdots+u^{(k)}(x)g_k^{(j)}(x)$$

用 $\boldsymbol{G}(x)$表示生成多项式矩阵，有

$$\boldsymbol{G}(x)=\begin{bmatrix}g_1^{(1)}(x) & g_1^{(2)}(x) & \cdots & g_1^{(n)}(x)\\ g_2^{(1)}(x) & g_2^{(2)}(x) & \cdots & g_2^{(n)}(x)\\ \vdots & \vdots & & \vdots\\ g_k^{(1)}(x) & g_k^{(2)}(x) & \cdots & g_k^{(n)}(x)\end{bmatrix}$$

输出码序列的相应码多项式为

$$\begin{aligned}c(x)&=\boldsymbol{u}(x)\cdot\boldsymbol{G}(x)\\&=[\,u^{(1)}(x)\quad u^{(2)}(x)\ \cdots\ u^{(k)}(x)]\cdot\boldsymbol{G}(x)\\&=[c^{(1)}(x)\quad c^{(2)}(x)\ \cdots\ c^{(n)}(x)]\\&=c^{(1)}(x^n)+x\,c^{(2)}(x^n)+x^2\,c^{(3)}(x^n)+\cdots+x^{n-1}c^{(n)}(x^n)\end{aligned}$$

【例 10.4】 将图 10-2 的（2，1，2）卷积码用多项式表示。

由前面的讨论可知，基本生成矩阵 $\boldsymbol{g}_\infty=[11\ 10\ 11\ 00\ \cdots]$，因而

$$\boldsymbol{g}^{(1)}=111\qquad g^{(1)}(x)=1+x+x^2$$
$$\boldsymbol{g}^{(2)}=101\qquad g^{(2)}(x)=1+x^2$$
$$\boldsymbol{G}(x)=[1+x+x^2\quad 1+x^2\,]$$

设 $\boldsymbol{u}=[10101]$，有

$$\begin{aligned}c(x)&=u(x)\cdot\boldsymbol{G}(x)\\&=[1+x^2+x^4]\,[1+x+x^2\quad 1+x^2\,]\\&=[1+x+x^3+x^5+x^6\quad 1+x^6]\\&=1+x^2+x^6+x^{10}+x^{12}+x+x^{13}\end{aligned}$$

与码序列[11 10 00 10 00 10 11 00 …]一致。

同理，由图 10-3 所示的（3，2，3）卷积码，有

$$\boldsymbol{G}(x)=\begin{bmatrix}1 & 0 & 1+x\\ 0 & 1 & x+x^2\end{bmatrix}$$

当 $\boldsymbol{u}=[10\ 11\ 01\ 11]$时，

$$\begin{aligned}c(x)&=\boldsymbol{u}(x)\cdot\boldsymbol{G}(x)\\&=\begin{bmatrix}1+x+x^3\\ x+x^2+x^3\end{bmatrix}\cdot\begin{bmatrix}1 & 0 & 1+x\\ 0 & 1 & x+x^2\end{bmatrix}\\&=[1+x+x^3\quad x+x^2+x^3\quad 1+x^3+x^4+x^5]\\&=[1+x^2+x^3+x^4+x^7+x^9+x^{10}+x^{11}+x^{14}+x^{17}]\end{aligned}$$

与码序列[101 110 010 111 001 001 000 …]一致。

10.3 卷积码的图形表示方法

在卷积码的概率译码中，图形描述是非常有用的。采用状态图、树图或网格图可以形象而准确地描述卷积码的编码和译码过程。

10.3.1 状态图

在卷积码编码器中，寄存器任一时刻存储的数据称为编码器的一个状态，随着信息序列的不断输入，编码器的状态在不断变化，同时输出的码元序列也相应地发生改变。

图 10-2 的（2，1，2）卷积码编码器由两级移位寄存器组成，因此状态只有 4 种可能：00，10，01，11，用符号 S_i 表示，分别将其对应为 S_0，S_1，S_2 和 S_3。表 10-3 和图 10-4 为（2，1，2）卷积码的状态表和状态图。

表 10-3 （2，1，2）卷积码状态表

输入 u	初态 S_i	次态 S_j	输出 C
0	00	00	00
1	00	10	11
0	10	01	10
1	10	11	01
0	01	00	11
1	01	10	00
0	11	01	01
1	11	11	10

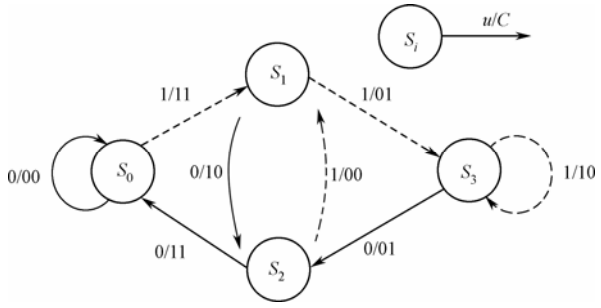


图 10-4 （2，1，2）卷积码状态图

在状态图中，用实线表示信息 0 输入，虚线表示 1 输入，若输入信息序列 $\boldsymbol{u}$ = [10101]，则状态转移为 $S_0 \rightarrow S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow S_2 \rightarrow S_0$，相应输出码元序列为[11 10 00 10 00 10 11]。编码器输出的码元序列是在信息序列的第一个码元输入直到最后一个码元完全移出移位寄存器过程中所产生的。这就要求有用信息序列输入完毕后，还应再向编码器输入 mk 个全 0 码元，所以最终状态应回到初始状态 S_0。

10.3.2 树图

如果（n，k，m）卷积码编码器的输入信息序列是半无限长序列，则它的输出码元序列也

应是半无限长序列，这种半无限长序列的输入、输出编码过程可用半无限码树图来表示。

图 10-5 示出了图 10-2 所示（2，1，2）卷积码的树图。

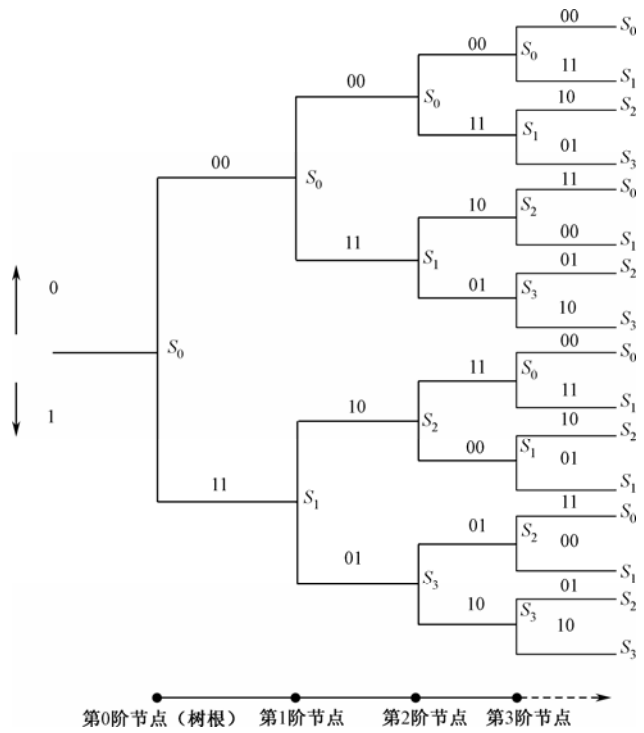


图 10-5　（2，1，2）卷积码树图

树图中，节点处符号为移存器状态，分支上的数据为输出码段，上分支为输入信息 0，下分支为输入信息 1。设移存器初始状态 $S_0 = 00$，当输入信息元 $u_0 = 0$ 时，由树根出发走上支路，移存器右移 1 位，状态仍为 S_0，输出码段 $\boldsymbol{c_0} = 00$；当 $u_0 = 1$ 时，则由树根出发走下支路，移存器状态转为 $S_1 = 10$，此时输出码段 $\boldsymbol{c_0} = 11$。在输入第 2 位信息元时，编码器已处于第 1 阶节点处，若在 S_0 点，则输入 $u_1 = 0$ 时走上分支，输出 $\boldsymbol{c_1} = 00$，新状态为 S_0；输入 $u_1 = 1$ 时走下分支，输出 $\boldsymbol{c_1} = 11$，新状态为 S_1；若在 S_1 点，则输入 $u_1 = 0$ 时，走上分支，输出 $\boldsymbol{c_1}$ =10，新状态为 $S_2 = 01$；输入 $u_1 = 1$，走下分支，输出 $\boldsymbol{c_1} = 01$，新状态为 $S_3 = 11$。再输入 u_2，编码器从第 2 阶节点处出发，此时起始状态有 4 种：S_0，S_1，S_2，S_3，按输入 0 走上分支，输入 1 走下分支的规则，得到相应的输出码段 $\boldsymbol{c_2}$ 和新状态。以此类推，输入无限长信息序列，就可以得到一个无限延伸的树状结构图。

从码树图上观察，输入不同的信息序列，编码器就走不同的路径，输出不同的码元序列。如输入信息序列 $\boldsymbol{u} = [10101]$，则输出码元序列 $\boldsymbol{c} = [11\ 10\ 00\ 10\ \cdots]$。

一般地，对于（n，k，m）卷积码来说，从每个节点发出 2^k 条分支，每条分支上标有 n 长输出数据，最多可有 2^{km} 种不同状态。状态图从状态上看最为简洁，但时序关系不清晰。码树图的最大特点是时序关系清晰，且对于每一个输入信息序列都有一个唯一的不重复的树枝结构相对应。它的主要缺点是进行到一定时序后，状态将产生重复且树图越来越复杂。

10.3.3 网格图

网格图又称篱笆图，它综合了状态图和树图的特点，结构简单，而且时序关系清晰。

从树图上观察，从某一阶节点开始所长出的分支从纵向看是周期重复的。例如，图 10-5 所示（2，1，2）码的树图中，当节点数大于 $m+1=3$ 时，状态 S_0，S_1，S_2，S_3 重复出现，因此在第 $m+1$ 阶节点以后，将树图上处于同一状态的同一节点折叠起来加以合并，就可以得到网格图。

图 10-6 所示为信息序列长度 $l=5$ 的（2，1，2）码网格图，实线表示输入为 0 的分支，虚线表示输入为 1 的分支，分支上标志的 n 位数据表示相应的编码输出 $\boldsymbol{C}$。从第 m 至 l 节点，编码器处于稳定的状态转移中，并且各节点的网格结构均相同；在 l 节点后 m 个移存器尚需转移 m 个状态，才能回到初始状态 S_0，由于 l 到 $l+m$ 节点过程中输入补充 0，所以只有实线分支。

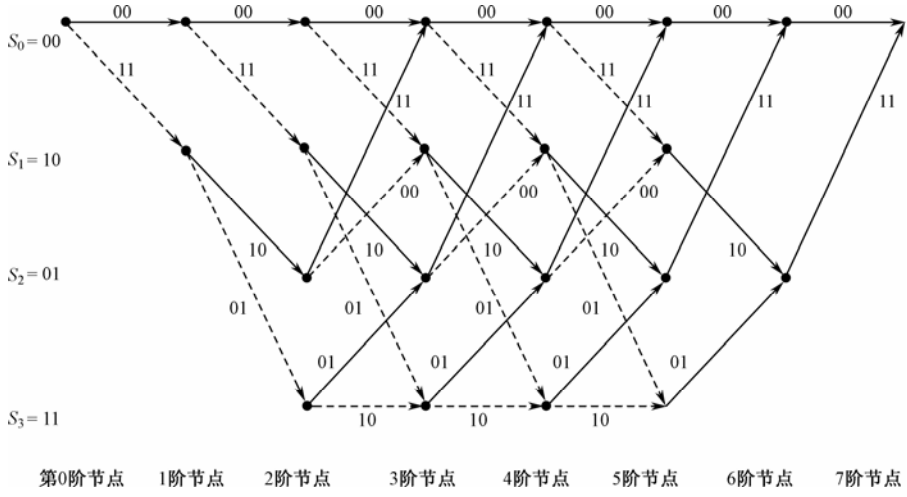


图 10-6 （2，1，2）卷积码网格图

与树图一样，网格图中每一种信息序列有唯一的网格编码路径，图中输入信息序列 $\boldsymbol{u}$ = [10101]路径对应的输出码元序列 $\boldsymbol{c}$ = [11 10 00 10 00 10 11]，与图 10-5 表示的结果一致。

10.4 Viterbi译码

卷积码有三种重要的译码方法：序列译码、门限译码和 Viterbi 译码。序列译码基于码的树图结构，能很好地处理约束度很长的卷积码，缺点是它的译码时间是可变的；门限译码基于码的代数结构，通过计算伴随式集合实现，缺点是在误码率方面表现较差；Viterbi 译码基于码的网格图结构，是一种极大似然译码方法。它具有以下优点：

① 有固定的译码时间；

② 适于译码器的硬件实现，运行速度快；

③ 译码的错误概率可以达到很小；

④ 容易实现，成本低。

10.4.1 Viterbi译码步骤

Viterbi 译码方法采用分段处理，每个码段根据接收的码元序列，按照极大似然译码准则，寻找发送端编码器在网格图上所经过的最佳路径，也就是在网格图上寻找与接收码相比差距最小的可行路径。对于 BSC 信道，这种寻找可等价为确定与接收码段具有最小汉明距离的路径。

对于（n，k，m）卷积码，假设已接收 l 个码段，Viterbi 译码算法可归纳为以下步骤。

① 在 $j = m$ 节点处，对进入每一状态的长度为 $j = m$ 的部分路径，计算部分路径的输出数据与对应接收的 j 个 n 长码段的汉明距离。将部分路径存储作为被留选的幸存路径。

② j 增加 1，把此时进入每一状态的所有分支（最多有 2^k 条）与前一阶节点处留选的幸存路径累积计算与相应接收码段的汉明距离，每个状态选出并存储其中一条汉明距离最小者，留选为对应的幸存路径，其余路径则删除。

③ 重复步骤②，直至 $j = l + m$，最终整个网格图中只剩一条幸存路径，译码结束。

由 m 阶节点至 l 阶节点，网格图中 2^{km} 个状态中每一个状态有一条幸存路径，但在 l 阶节点后，网格图中的状态数目减少，幸存路径随状态数相应减少，最后到 $l + m$ 阶节点时，仅剩一条幸存路径，这条路径就是要寻找的具有极大似然函数的路径，也就是译码器输出的最佳估值码元序列 $\hat{\boldsymbol{C}}$ 的路径。

如果在某阶节点时，某状态的两条路径具有相同的汉明距离，这时尚需观察下一阶节点累积的汉明距离，再选定最小距离的路径。

10.4.2 Viterbi译码

【例 10.5】 输入信息序列 $\boldsymbol{u}$ = [10101]至图 10-2 编码器，已知通过 BSC 送入译码器的序列 $\boldsymbol{y}$ = [11 10 01 11 00 10 11]，采用 Viterbi 译码算法对信息序列和码序列进行估值。

前面我们已推导出，当输入信息序列 $\boldsymbol{u}$ = [10101]时，正确的输出码序列是 $\boldsymbol{c}$ = [11 10 00 10 00 10 11]，与实际接收序列 $\boldsymbol{y}$ 比较有 2 个码元错误。根据图 10-6 的网格图，Viterbi 译码器对接收序列 $\boldsymbol{y}$ 的译码过程示于图 10-7 中，$\boldsymbol{y_i}$ 为接收码段，d 为汉明距离，$\hat{\boldsymbol{u}}$ 为信息估值。

在图 10-7（a）中，从初始状态到达 j = 2 阶节点的 4 种状态有 4 条路径，这 4 条路径与接收序列 y_0y_1 = [11 10]的汉明距离分别为 3，3，0，2，依次作为这 4 种状态的幸存路径。

当 j = 3 时，如图 10-7（b）所示，沿前一阶节点的幸存路径到达 S_0 状态有 2 条路径 $S_0 \xrightarrow{00} S_0 \xrightarrow{00} S_0 \xrightarrow{00} S_0$ 和 $S_0 \xrightarrow{11} S_1 \xrightarrow{10} S_2 \xrightarrow{11} S_0$，它们与 $y_0y_1y_2$ = [11 10 01]的汉明距离分别为 4 和 1，选取最小者[11 10 11]路径为 S_0 状态的幸存路径。同样 S_1，S_2，S_3 状态也都有 2 条路径，分别留选距离最小者为幸存路径，其余路径则删除。

接着由 j = 3 的 S_0，S_1，S_2 状态转移到 j = 4 的 S_0，S_1，S_2，S_3 状态，如图 10-7（c）所示，有 4 条路径与 $y_0y_1y_2\,y_3$ = [11 10 01 11]比较，具有最小汉明距离，被留选下来。

同样由 j = 4 到 j = 5 节点，如图 10-7（d）所示，每种状态都各自留选了一条幸存路径。

在图 10-7（e）中，当 j = 6 时，有用信息已输入完毕，输入端补充 0 至编码器，所以只剩下 S_0，S_2 两种状态，而 S_0 状态的 2 条路径与 $y_0y_1y_2\,y_3\,y_4\,y_5$ = [11 10 01 11 00 10]的距离都是 3，因此都被留存。

到 j = 7 时，，如图 10-7（f）所示，回到初始状态 S_0，只剩唯一的一条幸存路径，这条路径对应的输出码序列就是接收码序列的最佳估值 $\hat{\boldsymbol{y}}$ = [11 10 00 10 00 10 11]，相应的信息序列估

值 $\hat{\boldsymbol{u}}=[10101]$。

可见，译码结果与编码器的输出结果一致，实现了正确译码。

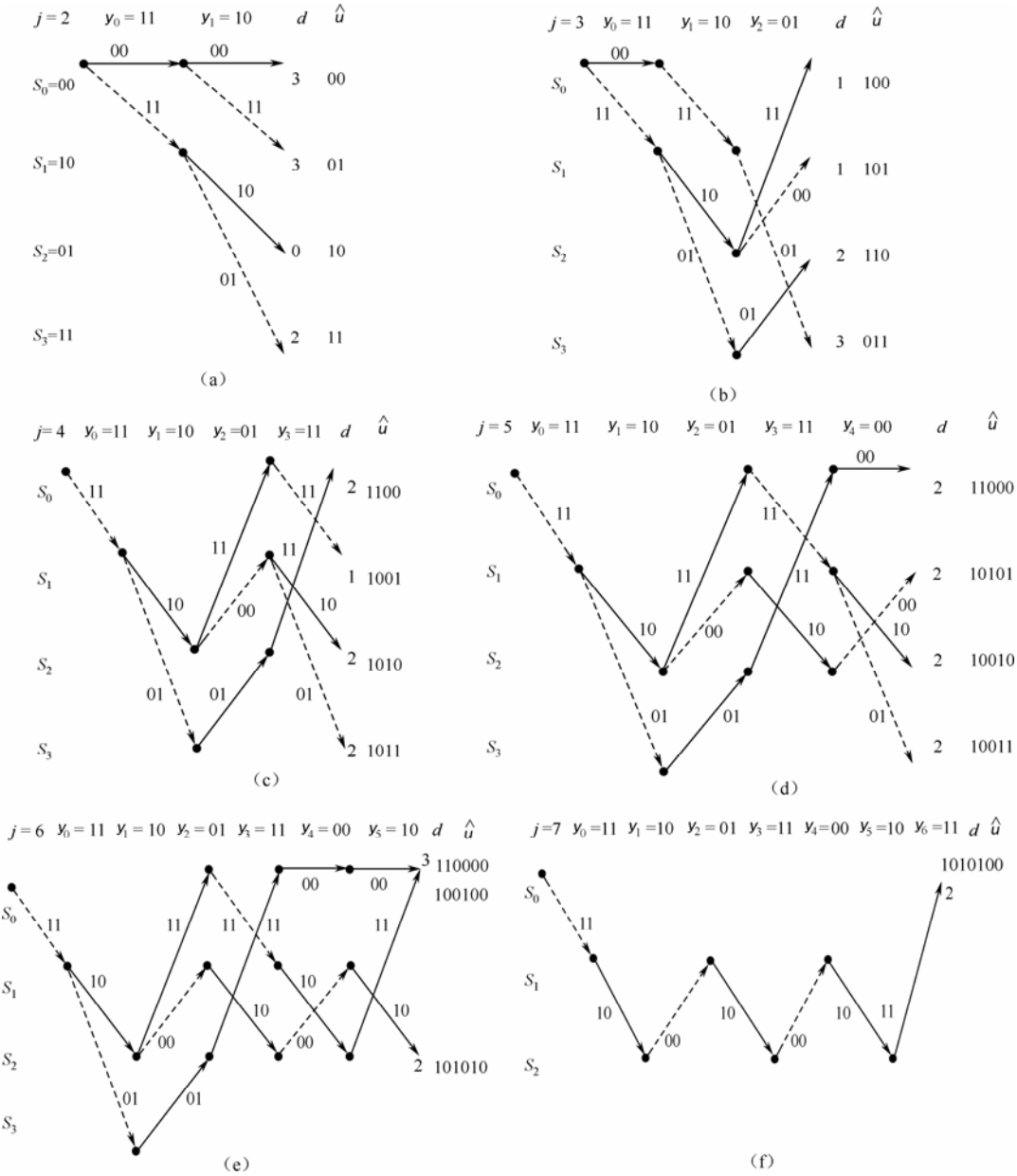


图 10-7　Viterbi 译码过程

本 章 小 结

卷积码利用了前后码段之间的相关性，纠错能力强于分组码，是一种重要的纠错码。本章介绍的主要内容有：

（1）卷积码的定义及主要参数。

（2）卷积码的生成矩阵和生成多项式描述。

（3）卷积码的状态图、树图和网格图描述。

（4）基于码的网格图结构的 Viterbi 译码算法的特点和步骤。

思考题与习题

10.1　什么是卷积码？编码存储和编码约束度的含义是什么？

10.2　卷积码的生成矩阵有何特点？生成矩阵为什么是半无限矩阵？

10.3　Viterbi 译码就是极大似然译码，这种说法对吗？

10.4　已知（2，1，3）码，生成序列 $\boldsymbol{g}^{(0)}=[1101]$，$\boldsymbol{g}^{(1)}=[1111]$。

（1）求出该码的生成矩阵 $\boldsymbol{G}_{\infty}$ 和生成多项式 $\boldsymbol{G}(x)$。

（2）求对应于信息序列 $\boldsymbol{u}=[11101]$的码序列。

（3）此码是否是系统码？

10.5　某（3，1，2）卷积码，生成序列有 $\boldsymbol{g}^{(0)}=[100]$，$\boldsymbol{g}^{(1)}=[101]$，$\boldsymbol{g}^{(2)}=[111]$。

（1）画出该码编码器。

（2）写出该码生成矩阵。

（3）若输入信息序列为[110101]，求输出的码序列是多少？

10.6　已知（3，1，2）卷积码，生成多项式有 $g^{(1)}(x)=1+x+x^2$，$g^{(2)}(x)=1+x$，$g^{(3)}(x)=1+x^2$。

（1）画出该码的状态图。

（2）画出 $l=4$ 的树图和网格图。

（3）用 ViterBi 译码算法对接收序列 101100001011111101 译码。

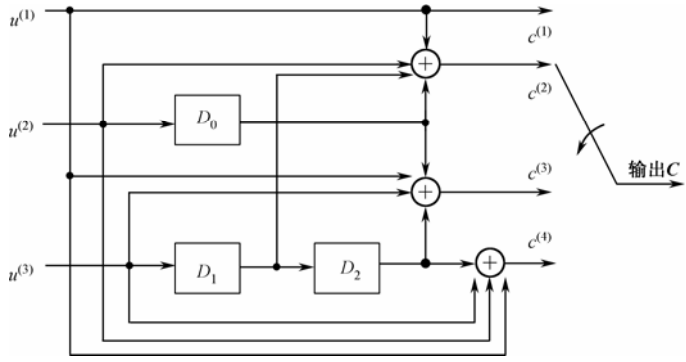
10.7　已知$(3,2,1)$卷积码，$g_1^{(1)}(x)=1$，$g_1^{(2)}(x)=x$，$g_1^{(3)}(x)=1+x$，$g_2^{(1)}(x)=x$，$g_2^{(2)}(x)=1$，$g_2^{(3)}(x)=1$，求 $u(x)=[1+x+x^3 \quad 1+x^2+x^3]$时的码多项式 $c(x)$。

10.8　某卷积码编码器如题图 10-1 所示。

（1）（n，k，m）=？

（2）$\boldsymbol{G}$=？ $\boldsymbol{G}(x)$=？

（3）$\boldsymbol{u}=[110\ 011\ 101]$ 时 $\boldsymbol{c}$=？



题图 10-1　卷积码编码器

10.9　已知$(2,1,2)$卷积码编码器的输出与信息位的关系为 $c^{(1)}=u_1+u_2$，$c^{(2)}=u_1+u_2+u_3$，当接收序列为 1000100000 时，试用 Viterbi 译码法求解发送信息序列。

附录A　GF（2^m）中元素的最小多项式和本原多项式（$1<m\leqslant 8$）

m =2			
1 (0,1,2)			
m =3			
1 (0,1,3)	3 (0,2,3)		
m = 4			
1 (0,1,4)	3 (0,1,2,3,4)	5 (0,1,2)	7 (0,3,4)
m = 5			
1 (0,2,5)	3 (0,2,3,4,5)	5 (0,1,2,4,5)	7 (0,1,2,3,5)
11 (0,1,3,4,5)	15 (0,3,5)		
m = 6			
1 (0,1,6)	3 (0,1,2,4,6)	5 (0,1,2,5,6)	7 (0,3,6)
9 (0,2,3)	11 (0,2,3,5,6)	13 (0,1,3,4,6)	15 (0,2,4,5,6)
21 (0,1,2)	23 (0,1,4,5,6)	27 (0,1,3)	31 (0,5,6)
m = 7			
1 (0,3,7)	3 (0,1,2,3,7)	5 (0,2,3,4,7)	7 (0,1,2,4,5,6,7)
9 (0,1,2,3,4,5,7)	11 (0,2,4,6,7)	13 (0,1,7)	15 (0,1,2,3,5,6,7)
19 (0,1,3,6,7)	21 (0,2,5,6,7)	23 (0,6,7)	27 (0,1,4,6,7)
29 (0,1,3,5,7)	31 (0,4,5,6,7)	43 (0,1,2,5,7)	47 (0,3,4,5,7)
55 (0,2,3,4,5,6,7)	63 (0,4,7)		
m = 8			
1 (0,2,3,4,8)	3 (0,1,2,4,5,6,8)	5 (0,1,4,5,6,7,8)	7 (0,3,5,6,8)
9 (0,2,3,4,5,7,8)	11 (0,1,2,5,6,7,8)	13 (0,1,3,5,8)	15 (0,1,2,4,6,7,8)
17 (0,1,4)	19 (0,2,5,6,8)	21 (0,1,3,7,8)	23 (0,1,5,6,8)
25 (0,1,3,4,8)	27 (0,1,2,3,4,5,8)	29 (0,2,3,7,8)	31 (0,2,3,5,8)
37 (0,1,2,3,4,6,8)	39 (0,3,4,5,6,7,8)	43 (0,1,6,7,8)	45 (0,3,4,5,8)
47 (0,3,5,7,8)	51 (0,1,2,3,4)	53 (0,1,2,7,8)	55 (0,4,5,7,8)
59 (0,2,3,6,8)	61 (0,1,2,3,6,7,8)	63 (0,2,3,4,6,7,8)	85 (0,1,2)
87 (0,1,5,7,8)	91 (0,2,4,5,6,7,8)	95 (0,1,2,3,4,7,8)	111 (0,1,3,4,5,6,8)
119 (0,3,4)	127 (0,4,5,6,8)		

表中括号中的值是最小多项式的幂次，例如，$m=6$ 下面的“3 (0,1,2,4,6)”表示 $m_3(x)=1+x+x^2+x^4+x^6$，由α^3（α 为本原元）的共轭根组构成；标下划线者为本原多项式。

附录B　熵函数计算用简明对数表

p	$-\log_2 p$	p	$-\log_2 p$	p	$-\log_2 p$	p	$-\log_2 p$	p	$-\log_2 p$
0.01	6.644	0.21	2.255	0.41	1.286	0.61	0.713	0.80	0.304
0.02	5.644	0.22	2.184	0.42	1.252	0.62	0.690	0.82	0.286
0.03	5.059	0.23	2.120	0.43	1.218	0.63	0.667	0.83	0.269
0.04	4.644	0.24	2.059	0.44	1.184	0.64	0.644	0.84	0.252
0.05	4.322	0.25	2	0.45	1.152	0.65	0.621	0.85	0.234
0.06	4.059	0.26	1.943	0.46	1.120	0.66	0.599	0.86	0.218
0.07	3.837	0.27	1.889	0.47	1.089	0.67	0.578	0.87	0.201
0.08	3.644	0.28	1.837	0.48	1.059	0.68	0.556	0.88	0.184
0.09	3.474	0.29	1.786	0.49	1.029	0.69	0.535	0.89	0.168
0.10	3.322	0.30	1.737	0.50	1	0.70	0.515	0.90	0.152
0.11	3.184	0.31	1.681	0.51	0.971	0.71	0.494	0.91	0.136
0.12	3.059	0.32	1.644	0.52	0.943	0.72	0.474	0.92	0.120
0.13	2.943	0.33	1.598	0.53	0.916	0.73	0.454	0.93	0.105
0.14	2.837	0.34	1.556	0.54	0.889	0.74	0.434	0.94	0.089
0.15	2.737	0.35	1.515	0.55	0.863	0.75	0.415	0.95	0.074
0.16	2.644	0.36	1.474	0.56	0.837	0.76	0.396	0.96	0.059
0.17	2.556	0.37	1.434	0.57	0.811	0.77	0.377	0.97	0.044
0.18	2.474	0.38	1.396	0.58	0.786	0.78	0.358	0.98	0.029
0.19	2.396	0.39	1.358	0.89	0.761	0.79	0.340	0.99	0.015
0.20	2.322	0.40	1.322	0.60	0.737	0.80	0.322	1.00	0

参 考 文 献

[1] Frank J．Cerra. Principles Of Digital Communication And Coding．Inc，McGraw-Hill，1979.
[2] Hanming R W．编码和信息理论．北京：科学出版社，1984.
[3] 王育民，梁传甲．信息与编码理论．陕西：西北电讯工程学院出版社，1986.
[4] 陈宗杰，左孝彪．纠错编码技术．北京：人民邮电出版社，1987.
[5] 吴伯修，归绍升．信息论与编码．北京：电子工业出版社，1987.
[6] 许树声．信号检测与估计．北京：国防工业出版社，1987.
[7] 归绍升．纠错编码技术和应用．上海：上海交通大学出版社，1988.
[8] 王新梅，肖国镇．纠错码——原理与方法．西安：西安电子科技大学出版社，1991.
[9] 方军，俞槐铨．信息论与编码．北京：电子工业出版社，1994.
[10] Guiasu S．Information Theory With Application．Inc．McGraw-Hill，1997.
[11] 胡冠章．应用近世代数（第 2 版）．北京：清华大学出版社，1999.
[12] 吴伟陵．信息处理与编码．北京：人民邮电出版社，1999.
[13] 刘颖. 数字通信原理与技术．北京：北京邮电大学出版社，1999.
[14] 张宗橙．纠错编码原理和应用．北京：电子工业出版社，2000.
[15] 王新梅，肖国镇．纠错码——原理与方法．西安：西安电子科技大学出版社，2001.
[16] 傅祖芸．信息论——基础理论与应用．北京：电子工业出版社，2001.
[17] 曹雪虹，张宗橙．信息论与编码．北京：北京邮电大学出版社，2001.
[18] 王新梅，肖国镇．纠错码——原理与方法．西安：西安电子科技大学出版社，2001.
[19] 周荫清．信息理论基础．北京．北京航空航天大学出版社，2002.
[20] 周航慈，孙丽华．信息技术基础．北京：北京航空航天大学出版社，2002.
[21] 陈运，周亮，陈新．信息论与编码．北京：电子工业出版社，2002.
[22] 孙丽华，陈荣伶．信息论与编码．南昌：江西科学技术出版社，2002.
[23] 牛少彰，刘吉佑．线性代数．北京：北京邮电大学出版社，2003.
[24] 仇佩亮．信息论与编码．北京：高等教育出版社，2003.
[25] 沈连丰，叶芝慧．信息与编码．北京：科学出版社，2004.
[26] 樊昌信．通信原理．北京：国防工业出版社，2004.
[27] 傅祖芸．信息理论与编码学习辅导及精选题解．北京：电子工业出版社，2004.
[28] 吕峰．信息理论与编码．北京：人民邮电出版社，2004.
[29] 孙丽华．信息论与纠错编码．北京：电子工业出版社，2005.
[30] 周炯槃，庞沁华．通信原理．北京：北京邮电大学出版社，2005.